

《高校历史照片档案智能编研系统》

项目研究报告

承担单位： 安徽大学档案馆 中国人民大学档案学院

项目负责人： 张斌 马仁杰

国家档案局
WWW.SAAC.GOV.CN

目录

1 绪论.....	1
1.1 项目概述.....	1
1.2 研究方法.....	2
1.3 研究内容要点.....	2
2 国内外研究综述.....	4
2.1 国内研究现状.....	4
2.2 国外研究现状.....	5
2.3 研究现状评述与发展趋势.....	5
3 高校记忆工程构建模式研究.....	6
3.1 基于 IPO 的记忆工程构建模式.....	6
3.2 记忆资源建设的输入.....	7
3.2.1 高校记忆资源的选择维度.....	7
3.2.2 高校记忆资源收集渠道与方法.....	8
3.2.3 高校记忆资源存储.....	9
3.3 记忆资源的组织与整合.....	10
3.3.1 高校记忆资源数字化.....	10
3.3.2 高校记忆资源标准化描述.....	10
3.3.3 记忆资源重组.....	14
3.3.4 主题地图构建.....	15
3.4 多维度记忆资源服务输出.....	16
3.4.1 记忆资源服务类型.....	16
3.4.2 记忆资源展示服务策略.....	18
4 基于 OAIS 的照片档案长久保存模式.....	19
4.1 OAIS 参考模型的目标.....	20
4.2 OAIS 环境.....	20
4.3 OAIS 功能模型.....	20
4.4 OAIS 信息包.....	22
4.5 信息包转换.....	23

4.6 基于 OAIS 的数字记忆资源数据处理全过程.....	24
5 面向数字照片的档案智能编研.....	25
5.1 档案编研的基本理论.....	25
5.1.1 基本概念.....	25
5.1.2 原则与内容.....	26
5.1.3 基本流程.....	26
5.2 照片编研的记忆价值分析.....	28
5.2.1 丰富记忆形式, 描绘客观事实.....	29
5.2.2 补校恢复记忆, 提供事实参考.....	29
5.2.3 串联古今记忆, 呼唤情感回归.....	29
5.2.4 多视角展现记忆风貌, 体现公众参与地位.....	29
5.3 智能化编研模式研究.....	30
5.3.1 数字档案编研的“记忆立方”.....	30
5.3.2 智能解析: 基于记忆点的智能化资源关联.....	32
5.3.3 基于时空维度的智能化编研.....	34
5.3.4 基于主题维度的智能化编研.....	35
5.3.5 基于场景维度的智能化编研.....	36
5.3.6 照片档案编研可视化设计方案.....	37
6 高校历史照片档案智能编研系统设计.....	41
6.1 平台控制台设计.....	41
6.2 资源库管理功能设计.....	42
6.2.1 提交模块功能设计.....	42
6.2.2 资源交接功能设计.....	52
6.2.3 资源管理功能设计.....	54
6.3 网站管理功能设计.....	56
6.3.1 内容管理功能设计.....	56
6.3.2 样式管理功能设计.....	71
6.4 系统管理功能设计.....	74

6.4.1 个人管理功能设计.....	74
6.4.2 系统日志功能设计.....	75
6.4.3 系统设置功能设计.....	76
7 研究展望.....	78
7.1 可能存在的困难与解决方案.....	78
7.2 未来研究方向.....	78
参考文献.....	80

1 绪论

1.1 项目概述

安徽大学始建于 1928 年，在陈独秀、胡适、蔡晓舟、刘文典等新文化运动健将和皖籍名流的倡议下成立，是安徽省近代高等教育的奠基者。学校历经安徽大学、安徽省立大学、省立安徽大学、国立安徽大学、安徽大学等发展阶段，2018 年 9 月 16 日，安徽大学迎来 90 周年华诞。

值此契机，安徽大学档案馆于 2018 年 1 月正式启动了“安徽大学数字记忆网（智慧档案馆一期工程）”建设项目，作为献礼安徽大学 90 周年生日的礼物。在“安徽大学数字记忆网（智慧档案馆一期工程）”建设过程中，应用“高校历史照片档案智能编研系统”开展了档案馆馆藏历史照片档案资源的搜集、整理加工以及开发利用工作，在校史档案资料征集、编纂和研究的基础上，进行外化展示服务。安徽大学档案馆利用馆藏历史照片档案开展多种形式的宣传教育活动，传播安徽大学悠久的办学历史，展现其独特的校园文化。该项目生动、全面、直观地展现了安徽大学九十载发展历程、历史变迁、缅怀先哲大师、记录学校各领域的发展成就和办学成果。

本项目由安徽大学档案馆与中国人民大学档案学院合作完成，安徽大学档案馆负责馆藏历史照片档案资源的收集、整理和网站展示内容的统筹规划，中国人民大学档案学院负责提供相关的技术支持和相关的理论研究工作。本项目自 2018 年初正式启动，经过双方共同努力，最终建成了安徽大学数字记忆网这一可视化成果，并于 2019 年 8 月 26 日完成了“安徽大学数字记忆网（智慧档案馆一期工程）”的验收，还通过了 2020 年安徽大学智慧档案馆二期工程建设方案。

为更好的提供系统构建与网站建设的理论基础与依据，本项目主要从高校记忆工程的构建方法入手，从理论层面进行梳理和构建，力求形成一套完整的历史照片档案资源开发利用模式与方法，为高校历史照片档案编研开辟一条新的路径。

本报告是对“安徽大学数字记忆网（智慧档案馆一期工程）”项目建设的成果总结，主要包括三个方面内容：高校记忆工程的构建、历史照片档案的智能编研、数字人文技术的应用。研究成果将作为本项目建设的主要理论依据，为

设计建设方法和路径提供支撑。

1.2 研究方法

(1) 文献调研法

充分利用安徽大学与中国人民大学的数据库资源，广泛收集和整理国内外与本项目相关的文献资源，进行文献研究，吸收已有研究成果，参考最佳实践案例，预测技术发展趋势。

(2) 实地调研法

对安徽大学档案馆进行充分的调研、访谈，重点针对照片档案的保存与利用问题进行需求调研，了解档案工作中的主要痛点与难点；对行业内档案管理、信息资源管理、信息技术等专家进行访谈，获取其对照片档案数字仓储与智能化利用方面的意见和建议；对照片档案资源进行调查分析，了解各地档案机构的照片档案资源情况，包括资源数量、资源载体格式、资源元数据方案等，为课题研究寻找事实依据。

(3) 数字人文研究方法

研究数字人文相关的信息技术，探究适合照片档案数字仓储与智能化利用相关的技术，如自然语言处理、人脸识别、照片优选、本体理论等，分析技术应用的结合点与应用场景，以为本研究提供技术支撑。

(4) 实践验证法

基于安徽大学档案馆馆藏历史照片档案资源，构建历史照片档案长久保存资源库管理平台，从照片档案的数字仓储，到照片档案的智能化利用，构建服务于全过程的系统平台。将其应用于建设“安徽大学数字记忆网（智慧档案馆一期工程）”，进而验证本课题的可行性与实践性，并从应用效果中分析验证本项目的科学性与适用性。

1.3 研究内容要点

高校历史照片档案智能编研系统的构建，目标在于解决安徽大学馆藏历史照片档案的管理与利用问题，解决档案馆在实际工作中的痛点，即传统历史照片档案资源量大无序，缺乏规范管理；档案编研技术陈旧，成果形式单一；档案展示空间受限，传播渠道狭窄。

具体研究内容要点如下：

(1) 历史照片档案的全生命周期管理

通过文献研究与实地访谈获取档案尤其是历史照片档案的全生命周期管理理论与实际需求。检索已有涉及档案、数字记忆资源全过程管理、全生命周期管理的文献，基于照片档案资源类型特性，对馆藏历史照片档案资源与征集历史照片档案资源的采集、加工、保管、利用环节进行总体设计。

(2) 历史照片档案的长久保存

通过文献研究获取档案尤其是历史照片档案的长久保存相关理论。主要聚焦于 OAIS（开放档案信息系统）的框架，包括其理论内涵、技术要点与具体应用案例，尝试构建基于 OAIS 的历史照片档案长久保存机制，构建照片档案长久保存资源库，以实现历史照片档案资源的标准化、系统化的长久保存。

(3) 历史照片档案数据的结构化组织

通过文献研究获取历史照片档案组织的一般模式与创新路径，基于数字人文方法，探索内容层面、数据层面的历史照片档案数据对象组织模式，实现历史照片档案数据的结构化组织，为下一步历史照片档案数据的智能化、自动化编研奠定基础。

(4) 历史照片档案数据的智能化编研

通过文献研究与系统工程研究方法，研究档案智能编研系统的构建，档案智能编研的模式等。具体技术要点包括，研究基于人工智能相关技术如何自动识别获取照片档案中的内容要素；历史照片档案数据的关联聚合的依据和维度划分；记忆性、故事性的线索的挖掘，内容发现模板设计与技术实现等，推动实现历史照片档案数据的成果输出，从而，提升编研效率与编研成果质量。

(5) 历史照片档案数据的可视化呈现

本项目启动时正值安徽大学九十周年校庆前夕，为了献礼安徽大学九十周年，安徽大学档案馆与中国人民大学档案学院合作，于 2018 年 1 月正式启动了“安徽大学数字记忆网（智慧档案馆一期工程）”项目。本项目建设的主要目的在于将安徽大学档案馆所保存的人才培养、科学研究、学科建设、文化传承、办学传统、服务社会、师资队伍建设、国际交流与合作、党建等方面多种类型和载体的校史档案资料以及校史馆所展示的校史照片档案资料，在网站中进行

呈现；同时，运用数字人文方法，对历史照片档案资源进行故事化开发，在“安徽大学数字记忆网（智慧档案馆一期工程）”中进行叙事可视化呈现。

2 国内外研究综述

2.1 国内研究现状

本项目重点研究照片档案全过程数字化管理方式。文献调研表明，近年来我国对于照片档案数字化管理的研究主要集中于以下几个方面：

（1）专属领域的照片档案数字化

包括高校照片档案的数字化、城建档案照片的数字化、航天照片档案的数字化、档案馆照片档案数字化等问题。

（2）照片档案数字化管理模式

刘萍^①提出照片档案的数字化管理需要解决的几个问题和发展趋势。她提出对照片档案进行数字化管理需要：数据录入、建立数据库；数字化管理照片档案的利用；数字化管理照片档案的安全性。郑志玥^②以“宁波光影图录”为例探讨了城市照片档案的利用实践与思考。

（3）照片档案数字化管理技术

邹亚劼^③研究了 OCR 技术在照片资源的应用技术与系统构建方法。胡湖^④探究了人脸识别技术在照片档案挖掘利用领域的技术架构。刘力超^⑤等人构建了北京城市记忆照片资源地理空间化组织平台，将分散的照片资源整合，引入 GIS 地理信息系统，实现照片资源与地理信息的映射，将照片所反映的人物、时间、地理聚焦在方寸之间。

（4）照片档案管理的问题与对策

照片档案作为一种记忆资源，学界对于它的研究多依托声像档案的研究而

^①刘萍.照片档案数字化管理初探[J].云南档案,2011(03):57-58.

^②郑志玥.城市照片档案的利用实践与思考——以“宁波光影图录”为例[J].艺术科技,2013,26(07):72.

^③邹亚劼.基于 OCR 的文档图片检测与信息提取系统的研究[D].哈尔滨工业大学,2015.

^④胡湖.基于人脸识别的照片档案挖掘利用架构[C].档案管理与利用——方法技术创新:国家档案局档案科学技术研究所,2013:287-291.

^⑤刘力超,陈慧迪,江健,王墨竹.北京城市记忆照片资源地理空间化组织平台构建[J].计算机系统应用,2017,26(08):1-8.



开展。王雪飞^①对城市记忆背景下的声像档案遗产现状进行了分析，他认为声像档案主要存在保存上的分散性、数量上的不均衡性、实践管理工作的不平衡性等问题。吴德^②等人对遂宁市城建档案馆声像档案的现状进行了分析，并提出了建设档案数据库的构想。随后介绍了遂宁市城建档案馆声像档案管理系统的概况，提出了系统建设中的问题和改进方法。

2.2 国外研究现状

国外对于照片档案的管理和利用研究已经深入到各个领域。其中，以照片文件在信息系统中运用并服务于其他领域的研究较多，如 IAN D.ROBERTSON^③讨论了医院照片档案的管理工作和信息系统的构建；Wang Yung-Cheng^④提出了照片文件在动物学习领域的应用技术；Yang G L^⑤使用图片档案和通信系统进行脑肿瘤研究等。

而在记忆工程领域，照片档案是重要的记忆资源，是各种“记忆工程”项目的管理重点。如“佛罗里达记忆”^⑥收录从16世纪起至今共202000多张照片，内容涵盖佛罗里达州人民的生活、工作、消遣等内容。该项目将照片档案作为在线资源展示，方便公众浏览和获取。“美国记忆”^⑦工程收集了超过900万件关于美国历史和文化的数字化资源，其中，照片资源是重要的组成部分。由此可见，照片档案在记忆工程构建中发挥着重要的作用，照片资源的资源数量庞大，并且能够有效反映人类的历史与文化面貌，是最为吸引人的一种记忆资源。

2.3 研究现状评述与发展趋势

^①王雪飞.城市记忆视域下声像遗产的现状研究[J].昭通学院学报,2016,38(04):119-121.

^②吴德,杨杰.城市记忆——声像档案数据库建设的思考[J].城建档案,2014(05):18-21.

^③IAN D. ROBERTSON, TRAVIS SAVERAID. HOSPITAL, RADIOLOGY, AND PICTURE ARCHIVING AND COMMUNICATION SYSTEMS[J]. Veterinary Radiology & Ultrasound, 2008.

^④Yung-Cheng, Wang, and,etc. Response to "Picture archive and communication system (PACS) for the detection of fish bone: An animal study"[J]. Otolaryngology Head & Neck Surgery, 2008.

^⑤ Yang G L,Tan Y F,Loh S C,Lim C C T. Neuroradiology imaging database: using picture archive and communication systems for brain tumour research.[J]. Singapore medical journal,2007,48(4).

^⑥佛罗里达州图书馆和档案馆. 佛罗里达记忆[EB/OL][2020-05-01]. <http://radio.floridamemory.com/>.

^⑦王健, 王小丹. “佛罗里达记忆”特色分析[J]. 中国档案, 2016(03):68-69.

^⑧国会图书馆. 美国记忆[EB/OL]. <https://memory.loc.gov/ammem/index.html>.

综上所述，国内外学者对于照片档案的研究已取得了一定成果。从研究内容和应用领域来看，对照片档案的利用呈现知识化服务、高层次服务、特定领域服务的特点，但仍存在以下不足：

(1) 照片档案全过程数字化管理研究仍有较大空间

对照片档案管理的研究虽然很多，但关于照片档案全过程、数字化管理模式与技术的研究仍然较少。尤其是新技术背景下，需要拓展照片档案的数字化保护与利用，更充分地挖掘照片档案的内容价值。

(2) 缺乏对照片档案内容挖掘方面的研究

照片档案的内容真实、直观、生动、丰富，价值巨大。但是现有研究较少涉及对照片档案的内容挖掘。同样，在实际的工作中，此方面的技术应用也相对较少。

(3) 缺乏照片档案智能化利用方面的研究

照片档案的服务利用模式还相对比较单一，可以结合“记忆工程”项目的开展，拓展对照片档案的利用与服务方式，建立智能化的照片档案利用体系。

(4) 针对照片档案的信息系统构建实践研究仍显不足

目前对于照片档案的信息系统构建实践还比较少，尤其是国内各地“记忆工程”项目中照片档案的展示与利用，以数字化方式实现的还很少。对于此类系统功能和技术的应用研究还尚显不足。

本项目从以上几个角度，弥补现有研究的欠缺之处，并提出系统完整的照片档案数字仓储与智能化利用方法体系。

3 高校记忆工程构建模式研究

3.1 基于 IPO 的记忆工程构建模式

“IPO 图”即输入（Input）-处理（Process）-输出（Output）图的简称，是 IBM 公司最早提出并用于描述过程的图形工具。IPO 图主要用于计算机和软件工程领域，但对记忆资源建设过程的研究有很强的适用性；作为描述过程的工具，IPO 图能够实现记忆资源建设的顶层设计，从宏观上进行全过程的统一规划和总体设计；以记忆资源作为输入对象，通过中间环节的处理能够清晰展现原始资料转化为记忆资源的过程；灵活性强，在三大过程的基础上可细分为若干个子过程或者小过程，能够清晰表达资源建设过程的模式、顺序和步骤。

从数据层面看，“安徽大学数字记忆网（智慧档案馆一期工程）”项目的建设即为针对安徽大学档案馆馆藏历史照片档案的建设，通过打通数字记忆资源的全生命周期流程，实现记忆资源的综合服务 and 利用。以 IPO 模型为工具，以“阶段-过程-结果”为线索，划分安徽大学数字记忆照片资源的输入、处理和输出三大阶段。如图 1 所示。

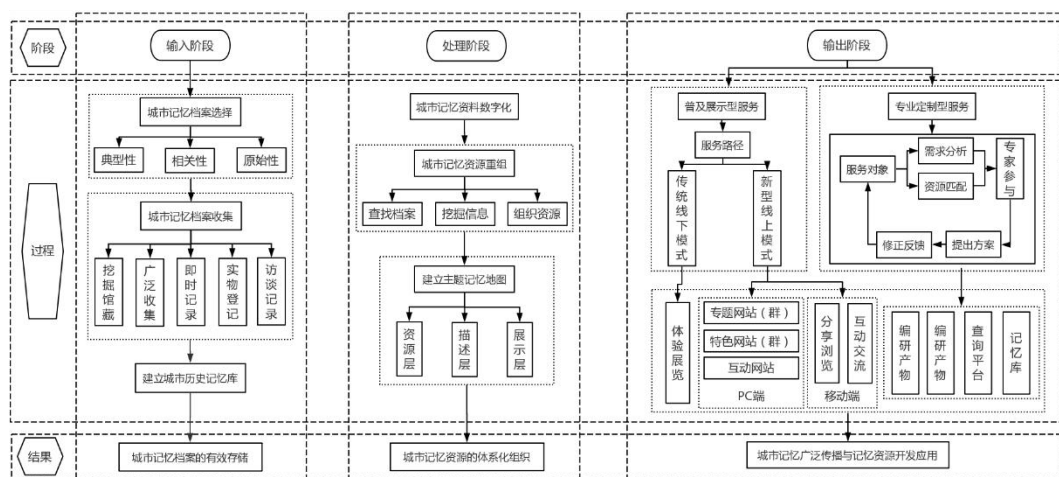


图 1 安徽大学数字记忆照片资源全过程管理

三大阶段过程不同，每个阶段都依赖于上个阶段的成果。输入阶段经过安大记忆档案资源的选择与收集并建立历史记忆库，最终实现记忆档案资料的有效存储；处理阶段通过记忆资源数字化、信息组织、构建记忆地图这一系列步骤实现记忆资源的体系化和结构化；输出阶段以两种形式的服务进行记忆资源的深度开发，最终实现人们对于历史记忆的广泛认知和全面应用，保护和传承记忆。

3.2 记忆资源建设的输入

记忆资源建设的输入过程实质是“选择—收集—保存”过程，即选取符合要求的原始档案，并依据档案来源、载体、形态的不同的采用针对性方法进行收集汇总并有效存储，为处理阶段奠定基础。

3.2.1 高校记忆资源的选择维度

高校记忆资料的选择是指按照一定的原则，采用科学、专业、高效的方法，对满足高校记忆要求的记忆资源进行甄别、鉴定与挑选。甄别选择符合要求的档案主要按照典型性、相关性和原始性三个维度。

(1) **典型性维度**。典型性维度是指档案所反映的内容对于所属领域、行业或主题来说有多大程度的代表性：十分典型，即档案信息能够代表某一校园、人物、事件等在某一时期、地域、领域范围所处的历史地位、发展状况，或是历史发展的关键节点，要集中力量进行搜集和抢救。非典型性，即档案所反映的内容属于主题范围，能够在一定程度上反映主题，但代表性不强，可选择一部分作为普遍性素材。无典型性，即不具有代表性的校园历史记录且相似的记录数量庞大，可暂且不予集中收集，在原单位保存即可。

(2) **相关性维度**。相关性维度是指记忆档案资料在多大程度上与所需主题有关联：直接相关，即直接描述主题，这类档案重点收集。间接相关，即不直接描述主题但内容与主题在事件、人物或时间上有很大关联的档案素材，也应收集归类。基本无关，即内容中未反映出与主题有关的任何信息或与主题关联性极小的档案，原则上不予考虑。

(3) **原始性维度**。原始性维度是指高校记忆档案中所反映和描述内容的加工程度：客观原真，即档案中的信息是对事件、人物等客观真实的描述，未经任何加工，是高校记忆建构最有价值的资源。如一份未经后人二次修改和加工的原始建筑设计图纸真实的表现了原始建筑风貌。转述引用，即档案是原始的，但信息是通过他人转述或加工的方式存在，可能带有一定的主观成分，但依然有较大的参考利用价值，要进行客观考量。如名人信件手稿等虽是原始记录，但信件中陈述的对象事实等带有写信者的立场态度和倾向性。主观臆造，即记忆材料是伪造、仿造的或经过篡改加工已无法反映主题内容，则基本没有收集价值，要通过科学考证避免记忆失真。

以三个维度为依据对数量庞大的馆藏历史照片档案进行鉴别和选择，挑选出具有典型代表性、主题相关性和客观原真性的档案作为记忆资料进行收集汇总，成为高校记忆建构的源泉。

3.2.2 高校记忆资源收集渠道与方法

高校记忆档案收集是依据不同来源和类型档案采用针对性的方法将资料汇总的过程，具体方式有如下五种。

(1) **挖掘馆藏**。档案馆被誉为“记忆的宫殿”，是保存人类记忆的各种表现形式、保存社会记忆、个人记忆的最权威场所”。珍藏大量反映校园发展各

方面情况的原始记录，其他馆藏机构如图书馆、博物馆、纪念馆等也会有高校记忆档案。这些机构的馆藏资源已经过专业鉴定，收集利用价值大且能节约时间和成本。要充分利用各馆专业的检索工具查找，还可通过馆际互借等方式获得。

(2) 广泛征集。校园中学院、机构部门、个人等产生的档案有很大一部分是由自身保存的，还有一些档案因历史原因散落教师职工手中。可通过线上线下的宣传方式鼓励个人、团体捐赠，可采用合理收购的方式，激发个人的捐赠热情，要积极联络校友、收藏家及馆藏机构，通过复制、交换等方式收集珍贵历史档案资料。

(3) 即时记录。校园整体及局部风貌是一个校园物化的历史形象，反映了校园空间及区域状态。尤其是经济发展迅速，校园面貌变化快。即时记录能够通过拍摄照片、视频的方式将校园及周边原样留存成为历史档案，根据不同时间的即时记录能够体现校园整体扩展迁移、局部改造变化情况。

(4) 实物登记。校园发展产生的记录并不都以纸质文件的形式保存，有许多记录或隐含在实物中的信息也是高校记忆档案的一部分。但这些实物因多种原因无法收集进馆，则采用拍照或登记方式加以记录，并进行相关保护工作。这种方式对于尤其是对实物遗产较多的校园十分适合。如老建筑、老牌匾、石碑石刻等都带有校园发展的鲜明印记，必须将其作为历史记忆留存下来。

(5) 访谈记录。访谈记录通过老教师、校友、事件亲历者等述说的方式重现历史，最终形成口述档案，能够弥补实体档案的不足，一些珍贵影像、事件、知识等也缺乏直接的记录，因此要通过访谈的形式获取。因访谈记录的主观性较强，要格外重视科学考证，保证真实性。

3.2.3 高校记忆资源存储

历史上产生的珍贵档案是中华民族历史文化的延续，是一所高校发展变迁的见证，是一个群体共有的精神财富。这些记录在保护和传承中华文明、塑造校园精神家园方面发挥不可替代的作用，必须加以有效保存。

(1) 按载体类型整理

收集所得高校记忆档案依据载体类型不同可大体分为文献类，主要通过文字、图形图表、各类符号等记录在纸质载体材料上，是高校记忆资源收集和保

存的主体部分；实物类，主要以实物的形式通过形状、材质、刻画符号等表现出独特的历史记忆信息；多媒体类，不同于传统纸质载体，主要依附于新型载体，如胶片、磁带等，需要特殊的保存条件。根据实际情况对这三类分别进行整理，按照规范的档案保存规章制度制定不同的归类及处理方案。

（2）建立校园历史记忆库

校园历史记忆库应该按照符合历史发展的逻辑顺序进行分类建库，以体现历史的延续性、全面性，将校园从产生、兴起到发展繁荣的历史脉络完整、清晰、系统地展现出来。可以从时间、机构、主题、大事件四个维度进行建库划分以作参考。这四种划分维度可交叉使用，以保证历史资源分类的穷尽性。

3.3 记忆资源的组织与整合

3.3.1 高校记忆资源数字化

数字化的目的就是要将不同形式、不同载体的原始档案以一定的标准格式进行预处理并以数字化的形式存储在统一的载体上，以方便资源组织整合和实现资源共享。包括将实体档案数字化和数字化档案标准化两个步骤。

（1）实体档案数字化

实体记忆档案资料保存在物理载体上，在输入阶段已将其进行归类。文字类可通过扫描技术、光学字符识别技术、人工录入等方式数字化。实物类可通过多角度拍照的方式、电脑绘图、三维影像模拟处理等方式数字化。声像类数字化则通过载体迁移，将不同载体（如胶卷，磁带，光盘等）上的记忆信息按照以一定的规范和要求迁移到统一的数字化载体上。

（2）数字化档案标准化

除已数字化的实体档案外，许多收集的档案本身就是数字化形式，如通过馆际互借的方式直接获得的电子版文件材料、声像影像资料，在网络中直接获取的文字、图片、视频资料等，存在来源结构各异、格式不统一的情况，不利于系统组织和整合。标准化的关键在于根据数据内容建立统一的著录规范和标引规则，研究数据格式标准、数据关联标准、信息编码规则等一系列规范，研究可信数字仓储、数字安全管理、长久保存技术等。

3.3.2 高校记忆资源标准化描述

当前，尚没有高校记忆资源元数据的相关标准，通过对各个领域已有元数据

标准的设计模式进行总结，分析高校记忆资源的元数据标准设计模式、设计思路与设计流程，对于节省元数据开发时间，提高开发效率，准确把握设计各环节有重要意义。

3.3.2.1 元数据标准的开发设计模式分析

(1) 全新开发设计模式

在物联网、古生物等专业性较强的领域，尚无完全成熟的元数据标准。这些领域专业性强，对元数据描述的准确性、著录的详尽程度、揭示能力、标准化程度要求非常高，因此需要全新开发设计以满足其高描述性的要求。

(2) 完全借用开发设计模式

殷沈琴^①等在分析电子教学参考资料的特征，根据元数据设计的通用原则，采用都柏林核心教育元数据 DC Education 描述简单对象，元数据编码及转换标准 METS 描述复杂对象，同时借用这两种元数据标准实现对电子教学资源的描述及共建共享。

(3) 继承再开发设计模式

元数据可以继承，是源于事物群体之间存在共性这一事实。例如音乐作品与影视作品虽然是不同的对象，但二者之间却存在许多相似甚至是相同的属性。例如名称、作者、播放时长等。目前大部分元数据标准的开发都是在复用现有元数据标准的基础上进行局部调整。

复用 DC 的开发设计。王仁武^②等在设计 e-GIRS 时复用了 DC 的部分元素，并且根据政务领域需求扩展了相应的元素。上海图书馆名人手稿数字图书馆课题组在设计名人手稿馆元数据方案时，以 DC-Lib 为基础的，选取其中的部分元素并根据资源特点扩展了 4 个个别元素（获得方式、捐赠项、书写人和载体形态附注）。^③杨剑^④等在设计侨批元数据方案时，复用了 DC 中的 12 个元数据，同时扩展了批款、批路、收藏信息 3 个侨批个性元素。

^①殷沈琴,张育超,葛家翔.电子教学参考资料的元数据方案的设计及编码[J].现代图书情报技术,2007(02):14-17.

^②王仁武,杨洪山,陈家训.电子政务信息资源元数据标准的设计与实现[J].情报资料工作,2007(04):52-55.

^③上海图书馆名人手稿数字图书馆课题组.名人手稿馆元数据方案的设计[J].图书馆杂志,2004,23(04):49-56

^④杨剑,杨明华,金文坚,刘晓莉.侨批元数据方案的设计和实现[J].图书情报工作,2013(03):100-104.

复用 CDWA 的开发设计。龚花萍^①等在数字文物博物馆的元数据设计中，选择了 CDWA 标准为依据，分别从文物基本信息、文物局部信息、文物相关历史文化脉络信息、文物相关视觉文献信息、文物展览信息和文物记录管理信息 6 个方面说明了数字文物博物馆文物信息资源的元数据模型。杨西宁^②等在对敦煌石窟艺术进行分析的基础上，参考 CDWA 元数据标准，形成了敦煌石窟壁画元数据标准、敦煌石窟彩塑元数据标准，并将两个标准成功应用于敦煌学数字图书馆项目中。

复用 ISO19115 的开发设计。国家基础地理信息中心、国图资源部信息中心等有关单位在采用 ISO19115:2003 的基础上编制了 TB/T19710-2005《地理信息元数据》。^③史建宗^④等参照地理信息国际元数据标准 ISO19115，以及生态、资源环境、气象和地质等元数据的国家标准，完成了冻土数据的扩展和冻土元数据标准的编制工作。

(4) 开发模式小结

全新开发设计模式适用于一些专业性要求较高且现有其他领域元数据标准的可继承内容极少的领域，这种元数据设计模式可以在本领域内发挥强大作用，但开发成本往往很高，普适性较差，因而很少使用。

完全借用开发设计模式与全新开发恰恰相反，即完全借用现有的通用元数据标准和相关领域的成熟元数据。这样虽然可以节省开发成本，但通用性较强的元数据标准难以满足高描述性的要求。

继承再开发设计模式通过选取相似程度较高的领域，继承其元数据集中通用的部分，开发专属于本领域的特色元素，既可以加快开发速度又可以节约开发成本。因此，对于高校记忆资源元数据的开发来说也可以采用这种模式。

3.3.2.2 元数据标准开发设计思路分析

高校记忆资源的来源、类型、格式多样，包括图书、期刊、手稿、电影、录音资料、照片等，资源的多样性必然导致信息组织过程中多种元数据标准的

^①龚花萍,孙晓,刘春年.文物信息资源元数据模型、实施标准与应用策略[J].情报科学,2013(02):80-84.

^②杨西宁,杜义涛,赵书城.敦煌石窟艺术元数据标准的设计及实现[J].上海交通大学学报,2008(91):221-225.

^③王杨刚,李娜,李玉龙,周丙锋,张敏,王小宁.境外地质矿产信息元数据标准与应用[J].地质通报,2015(08):1581-1588.

^④史健宗,南卓铜,赵林.多年冻土元数据标准研究和应用[J].遥感技术与应用,2014(05):878-885.

并存,为分布式信息环境下的信息资源整合利用带来了诸多问题和挑战:一方面不同类型的资源倾向于采用特定的元数据标准,另一方面用户又希望通过统一接口获取各类资源以满足其需求^①。因此,在开发高校记忆资源元数据标准时,应根据资源本身的特性选取相似度较高的领域,同时着重解决元数据互操作问题,进而开发出适用于高校记忆资源的元数据标准。

(1) 核心元素的设计思路

郑雯译^②等指出元数据元素的设置宜简不宜繁,在保证资源具有一定描述能力的前提下尽量简洁。李晓玲等^③在探讨服装元数据标准时提出要尽可能地复用标准方案、复用元素或复用修饰词及扩展方式,以及建立映射、转换机制等方式来达成互操作性。刘振华^④指出在易于用户掌握与应用的同时要实现对资源的准确描述,突显资源本身特性。上海图书馆名人手稿数字图书馆课题组指出元数据方案应尽可能采用精简的“核心”元素集以降低成本,同时满足对特殊资源的特殊描述要求。^⑤凌云等^⑥提出要针对每一种具体的资源类型选择适当的元数据标准,对特色资源进行准确和全面的描述。

(2) 面向用户需求的设计思路

李晓玲等^⑦认为元数据标准在结构与格式的设计、元素的增加与取舍等方面,应尽可能从用户的角度出发,为用户提供多层次的检索体系。杨剑等^⑧在侨批元数据方案的设计中也提出了终于原貌和符合用户现代检索习惯。刘振华^⑨在描述视频文件元数据方案的设计中提出要增加用户与系统之间的交互渠道,更好的满足用户需求。王仁武等^⑩指出用户需求是衡量元数据实用性的最终衡量标准,

^①宋琳琳,李海涛.大型文献数字化项目元数据互操作调查与启示[J].中国图书馆学报,2012(05):27-38

^②郑雯译,吴开华,赵阳.国外教育资源元数据标准比较研究[J].图书情报工作,2005(01):113-117.

^③李晓玲,黎焰.服装图片元数据标准初探[J].图书馆杂志,2008(06):63-65,34.

^④刘振华.视频文件元数据的设计与开发[D].山东大学,2009.

^⑤上海图书馆名人手稿数字图书馆课题组.名人手稿馆元数据方案的设计[J].图书馆杂志,2004,23(04):49-56

^⑥凌云,徐革,李一平,王利.特色数据库建设中的元数据标准选择——“峨眉山自然遗产特色数据库”的应用实例[J].情报杂志,2006(01):131-133.

^⑦李晓玲,黎焰.服装图片元数据标准初探[J].图书馆杂志,2008(06):63-65,34

^⑧杨剑,杨明华,金文坚,刘晓莉.侨批元数据方案的设计和实现[J].图书情报工作,2013(5):100-104.

^⑨视频文件元数据的设计与开发[D].刘振华.山东大学 2009

^⑩王仁武,杨洪山,陈家训.电子政务信息资源元数据标准的设计与实现[J].情报资料工作,2007(04):52-55.

因此要增加系统与用户之间的交互式对话能力。

(3) 可扩展的设计思路

王小丽等^①指出为了适应不同档案信息资源的需要,反映档案信息在不同生命周期的真实性,要特别注重元数据标准体系的可扩展性。龚花萍^②认为元数据模型的设计应具有良好的伸缩性,针对不同类型用户进行扩展。金更达^③提出整个元数据体系和每个元数据模块都应该可以扩展,保留细化元数据元素的空间以适应未来需求的变化。

由此可见,在设计高校记忆的元数据标准时,应首先从用户需求出发,尽可能的复用已有元数据标准中的元素,必要时针对高校记忆资源的特征设计特色元素,同时注意描述的深度,确保元数据标准的适用性和灵活性。

3.3.2.3 元数据标准的开发设计流程分析

综合各学者的观点,元数据的设计流程应包括以下几个环节:资源分析、建立模型、属性提取和文献调研、元素精炼、规范控制、确定限定规则和著录规则,最终形成标记方案,在具体的应用环境中还应关注制定元数据应用纲要,创建元数据记录/质量控制等问题,高校记忆资源的元数据标准的设计应包括以下几个主要环节:

①资源分析。对高校记忆资源进行分析,详细了解历史音像、图片资料、实物建筑等资源,明确高校记忆资源之间的相互关系,建立模型;

②属性提取和元素精炼。由高校记忆方面的专家从利用的角度提出属性要求,再经元数据专家对元数据标准规范进行分析,选择适当的设计模式和最为相关的元数据规范及研究成果,初步确定需要复用的元素和元素修饰词;

③确定标记方案。参照业界成熟的元数据标准规范对每一元素的著录内容和取值范围进行必要的限定,制定详细的著录规则并形成相应的规范文档,选择合适的标记方案,形成完整的高校记忆资源元数据标准。

3.3.3 记忆资源重组

(1) 查找档案资料。查找档案资料即确定查找的主题、关键词、时间、地点、作者等信息并通过有效方式找到与之相关的多份档案。可通过目录数据库、

^①王小丽,王芳.国内外数字档案馆元数据标准体系比较研究[J].情报科学,2007(03):382-389.

^②龚花萍,孙晓,刘春年.文物信息资源元数据模型、实施标准与应用策略[J].情报科学,2015(02):80-84.

^③金更达,何嘉荪.电子文件元数据标准设计框架研究[J].档案与建设,2005(04):4-7.

全文数据库及多媒体数据库等大型数据库中输入关键词等信息进行直接检索，也可利用传统的目录索引或通过翻阅图书、报纸、杂志等进行人工查找。检索要依据主题相关性，如人物相关、地域相关、事实相关等。对不确定的资料不能盲目选用，要进行查考检验，确定真实后再利用。

(2) 挖掘记忆信息。挖掘记忆信息即从档案资料中抽取有价值、符合主题要求的内容。要注意把握线索和关键，重要线索如时间关联、事件发展演变、人物关系等，关键节点如重要时间、事态转折、人物命运转折等，这些构成了历史记忆的“脉络”；要注意相关性，一份档案素材或一篇文章内容可能反映多方面信息，只需挑选与要求相关的，其余不需要深究；要注意照片图纸相关的文字说明中有价值的信息，图片中的隐含信息要用文字转述表达出来，图片没有的内容不能臆造。

(3) 组织记忆资源。组织记忆资源即将从档案中挖掘的记忆信息按历史逻辑关联提取、精炼、重组，实现对历史对象的描绘与表达，要遵循三点：一是真实反映。对记忆资源的提炼带有一定的主观性，但不能有任何的歪曲和臆造。二是历史逻辑。要按照历史逻辑顺序对记忆信息进行组织，不可颠倒顺序。三是标注说明。对于重要图片、绘图等素材可直接加以组织，但要进行客观标注和说明。

3.3.4 主题地图构建

主题记忆地图是以主题为依据实现记忆关联的三层空间模型，如下图所示。记忆地图能够将记忆资源结构化、体系化，建立历史记忆关联，并通过可视化方式展现，便于人们理解、认知和记忆。主题记忆地图可实现两个重要功能：一是导航功能，利用者可利用记忆链接通过目录层级准确定位记忆资源。二是关系揭示，通过特定关联可以揭示记忆资源之间并列、从属、相关等关系。利用者不仅可以找到所需要的记忆资源，也可通过关联链接找到其他相关资源。

国家档案局
www.saac.gov.cn

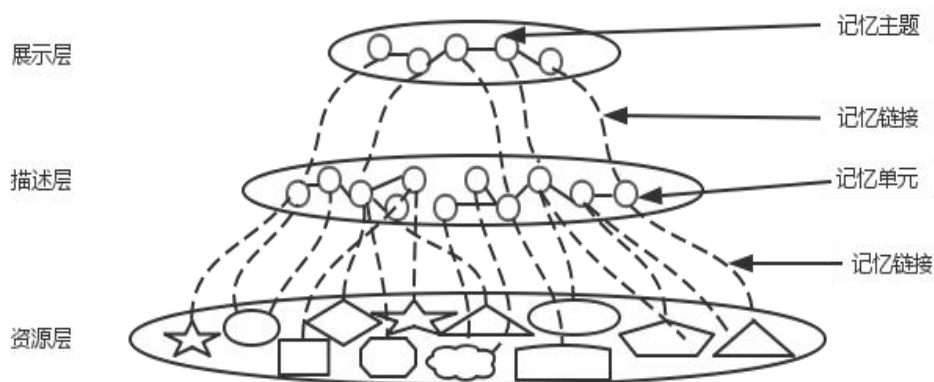


图 2 高校记忆地图模型图

(1) **记忆展示层**。该层通过可视化工具将记忆资源及其之间的关系映射出来，以空间网状地图的形式简单明了地进行展示。在人们的头脑中有许多符合人们认知的有价值的历史“记忆点”，我们将这些“记忆点”称之为“主题”，定义为“高校记忆资源中基于人类头脑认知的可被识记、保持和再现的单个对象”，该对象可以是一个学院、一位历史名人、一个事件，如眼镜湖、刘文典、档案馆等。属于同一类的主题可组成一个专题，如安徽大学的一个学院。

(2) **记忆描述层**。该层为中间层，上下关联，是多个记忆单元的集合。记忆单元以主题为依据，是关于记忆主题本身或记忆主题局部或与记忆主题对象有密切关联的一个描述单位，多个记忆单元以一定的逻辑联系串联起来组成了记忆主题的全部内容。

(3) **记忆资源层**。记忆资源层为最底层，包含了处理阶段所有从原始记忆素材提炼整合的记忆资源。多份记忆资源通过一定的相关性组合成了一个记忆单元。

3.4 多维度记忆资源服务输出

3.4.1 记忆资源服务类型

(1) 普及展示型服务

该服务以普及历史文化、传播高校记忆为宗旨，以覆盖率为优先考量，使高校记忆资源能够被最大程度地理解和认同，建构最广泛的高校记忆。具有普及性、全面性和教育性的特点。

国家档案局
www.saac.gov.cn

①普及展示型服务是一种受众广泛的主动性服务，要最大限度地拓展受众面，可通过线下和线上两种路径来实现高校记忆资源的输出。传统线下模式以传统媒介为依托，如报纸、广播、电视、出版物等。可在人员密集的地方、重大节庆日等举办高校记忆展览，如充分利用安徽大学档案馆等文化场所，开展主题记忆展览；在校园主要场所举办文化记忆宣传活动。可进行编研出版、制作记忆影片、开办历史文化讲座等形式进行多样化普及与展示。新兴线上模式则以现代化社交媒介为依托，如互联网和移动客户端等。充分利用网络和移动客户端不受时空束缚的特点，将常识性、科普性、知识性、文化性内容推送给大众。可建立高校记忆网站、开发移动客户端、创建高校记忆微信公众号，并在其中建立大众参与的互动平台，使人人都成为记忆创建者。

②展示内容。普及展示型服务面向广大群众，其内容要通俗易懂、贴近生活、趣味性强。主要提供原始档案资料如图纸、手稿、观测记录、合同、票证、书信、照片、视频等，是历史最可靠的凭证，是市大众了解历史最直接的方式；粗加工产品，专业人员通过对档案资源进行组织、整理、编研等实现对校园历史系统全面梳理。输出内容具有介绍性和知识性，使大众头脑中的记忆认知由点到线，由线及面。

（2）专业定制型服务

专业定制型服务以需求为导向实现资源深度开发和定向输出，旨在将高校记忆资源实现全面应用，推动校园发展。该服务主要面向机构、部门的档案管理需求、高校和科研机构的科学研究需求等，按照特定流程实现服务利用。

①需求导向的服务流程设计。专业定制型服务并不是撒网式的单线输出，而是“输出—接收—反馈”的循环模式，该服务主要包括五个步骤：第一步，需求分析。与用户进行深入沟通，明确用户想要达到的目的和效果，并对需求进行拆分，提取需求关键信息。第二步，资源匹配。按需求分析的结果提取关键词，利用高校记忆地图的导航功能、资源关联功能和可视化展示功能找到资源，为研究方案的制定提供素材。第三步，专家参与。通过专家网络邀请行业内权威的专业人员，采用科学方法和专业技术对相关高校记忆资源进行分析和挖掘，保证解决方案的权威性和可信度。第四步，提出方案。根据用户需求、利用高校记忆资源并进行查考、调研、分析、总结后，得出符合要求的解决问

题方案，如一份研究规划报告、行业预测或专业库。第五步，反馈修正。提出解决方案后，用户会在实践中反馈使用情况或者在此基础上提出新的需求，要对输出内容、方式、产品等进行修正以不断适应多方位需求。

②定制型服务成果。专业定制型服务从用户需求出发，具有高匹配度、科学权威和动态及时的特点。研究报告，通过对原始材料的挖掘与整合撰写分析报告、规划方案等，帮助用户了解发展沿革、分析利弊，提供参考；建立专门库，这是与用户进行长期深度合作的方式，如面向高校建立历史文献库。

3.4.2 记忆资源展示服务策略

面向公众的记忆资源展示是记忆资源服务输出中最为重要、核心的部分，是将一个高校记忆传播、传承下去的重要手段，只有得到公众认知、认可、认定的记忆才是发挥出其内在价值的记忆。因此，针对记忆资源的展示要求多维度、多层次、多途径、多方法。

3.4.2.1 多维度记忆资源展示

记忆资源具有多维属性，分类是多维的，自然也应该是多维展示的。对应分类维度，记忆资源可以按载体维度、主题维度、时空维度等进行展示。

载体维度：指按照记忆资源的数据类型分化聚类而成的各类资源，具体包括文本、音频、视频、图片、3D 模型等等。以某一载体类型的资源为中心，通过记忆线索的串连，形成独具特色的资源展示维度。其展示形式包括故事图集、口述史记忆等。

主题维度：以记忆地图为基本参考，将记忆资源按照展示内容的主题性分层次分类别进行归纳聚类，形成主题维度。主题维度的划分能够帮助高校记忆以某一特定主题内容为核心进行呈现。其展现形式包括专题网站和轻专题展示等。

时空维度：打破资源的二维限制，根据记忆资源的产生时间（或所描述的时间）和产生地点（或所描述的地点），利用 GIS 地理信息系统，将记忆资源分布到地图上——此为“空间”维度；利用时间线等工具，将记忆资源串联起来，展现不同时期的记忆资源——此为“时间”维度。“时间+空间”促成时空维度，用以展现安大记忆在某一特定时间、特定地点下的状态。具体展示形式有记忆地图等。

3.4.2.2 多层次记忆资源展示

记忆专题具有层次性。一方面，总体架构基于点、线、面对安大记忆资源进行解构，形成“资源+脉络”的数字模型；另一方面，大主题下有很多子主题，具有父类包含关系，例如安大记忆的一级分类可以是安大故事、历史建筑、学校、人物、中西交流、礼俗等等，其中学校的下一层次可以包括书院、乡学等。

通过深层挖掘记忆信息、组织记忆资源、绘制主题记忆地图，进而依据记忆地图构建网站展示体系。记忆地图在这里有两种含义，一种是概念上的，一种是空间上的。概念上的主题记忆地图以“记忆数据-数据单元-资源关联组织-资源服务”的流程进行组织展示；空间上的主题记忆地图则发挥地图的层级比例优势，由点及面，层层递进，打开资源全景世界，展现记忆资源全貌。

3.4.2.3 多途径记忆资源展示

线上平台线下展览相结合。一方面，借助于互联网技术，搭建数字记忆资源服务平台，打造精品展示网站，如专题网站、轻专题、特色网站、互动网站等；另一方面，不忽视线下展览的作用，同时面向资源与用户组织基本陈列与专题展览，地点不局限于档案馆，还可深入社区、校园等，通过学生志愿者进行传播与宣传教育；还可举办相关路演，提高平台与项目知名度。此外，利用移动互联网发展的极大机遇，积极拓展移动端的资源展示，开发 APP 或小程序应用。

3.4.2.4 多方法记忆资源展示

新兴技术发展迅猛，层出不穷。运用多种资源展示技术如 VR、AR、故事图集、今昔对比、多媒体技术、可视化技术等对资源进行呈现，可以提高记忆资源的展陈效果，让用户有身临其境之感，加强用户体验。同时新兴技术的应用的本身也是一种展示，强强联合，打造炫酷科技与人文情怀的双重享受。

4 基于 OAIS 的照片档案长久保存模式

OAIS，即开放档案信息系统（Open Archival Information System）的简称，一项旨在为基于长期保存目的的信息系统建立一个参考模型和基本概念框架，以维护信息系统中数字信息的长期保护和可存取。OAIS 是由国际空间数据系统咨询委员会（CCSDS）制定的标准，2003 年最终作为 ISO 的标准（ISO 14721:2003）颁发。OAIS(Open Archival Information System)是一个由人和系统组成的,承担信

息长期保存并将信息提供给目标团体(Designated Community)责任的存档机构。

4.1 OAIS 参考模型的目标

- 对长期保存处理数字信息所需理解的存档概念提供了一个框架；
- 为非存档组织有效参与保存工作提供所需概念；
- 为描述和比较结构、对现存和将来的存档机构操作提供包括命名和概念在内的框架；
- 为所保存数字信息的数据模型进行比较，为讨论数据模型和内在信息随时间如何变化提供基础；
- 为其它长期保存的非数字形式信息扩展提供一个基础；
- 为鉴别 OAIS 相关标准的产品提供向导。

4.2 OAIS 环境

OAIS 认为，OAIS 系统是一个介于生产者、消费者和管理者之间的存档系统。如下图所示，生产者 (Producer) 是提供信息需要保存的个人、机构和系统；管理者 (Manager) 是制定 OAIS 总体政策的实体，而这些政策将成为更广泛背景下政策领域的有机组成部分，它并不涉及日常的档案管理活动（具体的档案管理活动有 Administration 模块来实现）；信息用户 (Consumer) 是需要查找和获取 OAIS 所保存信息的个人、机构和系统。

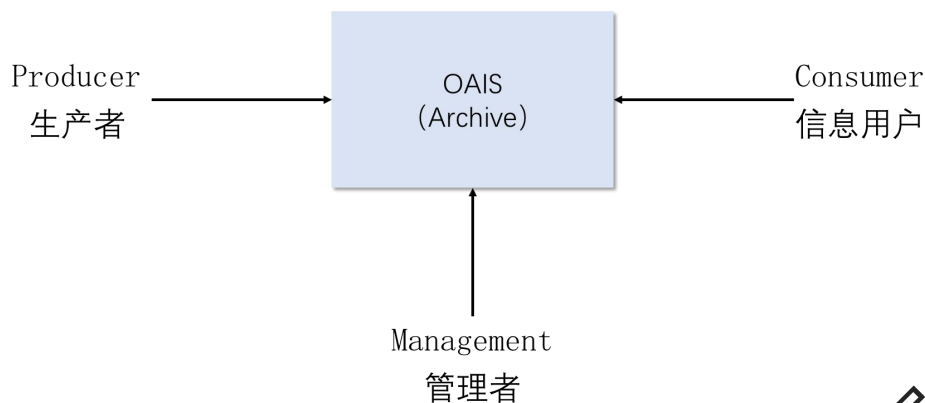


图 3 OAIS 环境

4.3 OAIS 功能模型

OAIS 系统包括以下 6 个功能模块:

国家档案局
WWW.SAAC.GOV.CN

- **摄取(Ingest):** 从信息生产者那里接收或收集按照一定格式组织的信息(SIP)。这些信息经过检验后建立相应元数据,元数据交给数据管理模块,信息被转换成长期保存规定格式组织的AIP,然后传递给长期存储模块保存。
- **长期存储(Archival Storage):** 实际存储AIP、负责建立具体存储与存取系统,并在获取功能模块提出请求时将AIP提供给该模块。
- **数据管理(Data Management):** 存储关于数字信息单元的元数据和关于长期保护处理政策、程序、技术和系统的元数据,并提供对这些元数据的检索与管理。
- **获取(Access):** 提供用户检索元数据和索取数字信息单元的界面,提供检索机制,并将AIP转换为适合用户利用的发布信息(DIP),还可能承担身份认证和授权管理责任等。
- **管理(Administration):** 通过有关政策、规范、程序、 workflow 等来监测和控制整个长期保存系统的运行和各个模块的运行。
- **保存规划(Preservation Planning):** 监控OAIS环境,并为保证在技术过时等环境下OAIS系统仍长期可用而提供建议的服务和功能模块。具体包括评估存档内容和定期建议存档信息迁移、就存档机构标准和政策提出建议、监控技术发展及用户服务需求和知识储备的变化、制定具体的迁移计划、软件原型、测试计划以保证管理迁移目标的实施等。

除上述功能外,系统还提供各种各样的通用服务(Common Services),如操作系统服务、网络服务、安全服务等。因这些服务是一般信息系统普遍具有的,所以功能模型图中略。OAIS参考模型中各个功能的要素、要素间的关系、功能的具体实现等如下图所示。

国家档案局
www.saac.gov.cn

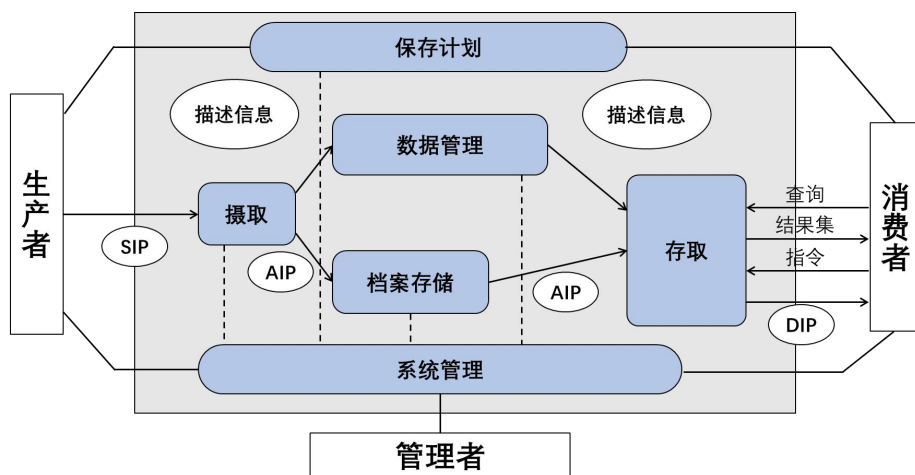


图 4 OAIS 参考模型

4.4 OAIS 信息包

信息包(Information Package)是 OAIS 信息模型中的核心概念，也是功能模型中各模块之间传递的基本对象。从信息包的构成来看，它包括：

内容信息(Content Information)： 是被保存的原始信息对象。

保存描述信息(Preservation Description Information, 下简称 PDI)： 描述内容信息的特征并使其得以完全保存的必要信息。具体包括来源信息(Provenance Information)、背景信息(Context Information)、参考信息(Reference Information)、固化信息(Fixity Information)。

打包信息(Packaging Information)： 将内容信息和保存描述信息联系、封装和识别的信息。内容信息、保存描述信息通过打包信息构成一个信息包,三者绑定成一个整体。

描述信息(Descriptive Information)： 是描述信息包的特性和属性并帮助定位和检索用户所需信息包的信息。它们之间的关系如下图所示：

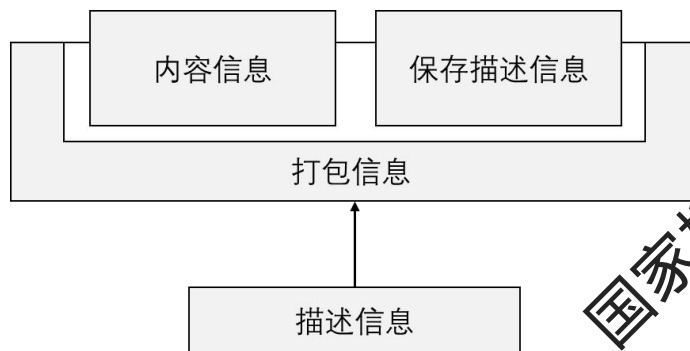


图 5 信息包结构

国家档案局
www.saac.gov.cn

从信息包的种类来看，它可以分为三大类：

提交信息包(Submission Information Package,简称 SIP)：由信息生产者提交给 OAIS 的信息包，它的形式和具体内容通常由信息生产者和 OAIS 协商决定。

存档信息包(Archival Information Package,简称 AIP)：由 OAIS 保存的符合永久或长期保存属性的信息包。

发布信息包(Dissemination Information Package,简称 DIP)：OAIS 基于信息用户需求请求而传递给用户，从一个或若干个 AIP 中提取出来的信息包。

SIP 又分为存档信息单元(Archival Information Unit, AIU)和存档信息集合(Archival Information Collection, AIC)两类。SIP、AIP、DIP 共同描述了信息在摄取、保存、管理、获取整个过程中的变化过程，其中 AIP 是长期保存的关键和中心点，使 OAIS 完成了保存和存取的责任。在 OAIS 参考模型中，对 AIP 及其关键组成和类别进行了详细的说明。

4.5 信息包转换

(1) 信息生产者实体中的数据转换：

首先当生产者做出将信息存储在 OAIS 的决定后， he 需要与档案工作者就 SIP 的内容、格式、储存时间等问题达成提交协议。在信息生产者和 OAIS 之间,SIP 定期向 OAIS 提交数据。尽管从物理上来说，描述对象和元数据既可以包括在数据对象中，也可以分散在许多独立的描述对象中。但是从逻辑上来说，数据提交过程必须被看成数据内容对象和描述对象的集合。

(2) 摄取功能模块中的数据转换：

OAIS 不需要保持 SIP 中的信息与原来完全一致，SIP 在 OAIS 中的形式和内容都可能发生变化。因此 SIP 与 AIP 的映射关系并不是一一对应的。摄取过程首先将从数据接收环节得到的 SIP 转换成能为长期存储和数据管理功能模块接收的 AIP 和包描述。摄取过程的复杂程度因不同的 OAIS 和信息生产者而异；其次，它将接收来的信息对象分类，判断它们属于现存的哪个集合。在 AIP 中作为档案存储时产生相关集合描述更新的提示信息。最后，摄取功能协调数据管理模块与长期存储模块之间的更新，同时适时提供协调和错误恢复。

(3) 长期存储和数据管理功能模块中的数据转换：

长期存储功能模块将 AIP 转成长久保存的档案信息，数据管理功能模块接收由摄取功能模块产生的包描述并且补充现存的集合描述以囊括新接收的包描述信息。在这一过程中，OAIS 有责任维护软件拷贝的可用性或者通过内部格式的仔细记录以保证将来数据能够转移到其他系统中而不造成信息丢失。

(4) 获取功能模块中的数据转换：

当信息用户者需要使用信息时，检索工具能够提供给用户关于 OAIS 中储存所有信息的全景，帮助用户定位查找信息；当用户确认需要查找 OAIS 中的信息时，用户使用请求工具提出获取数据请求；获取功能模块在数据管理模块中记录下用户的请求并确认是否能够满足，并且与长期存储和数据管理模块联系请求 AIP 和相关的包描述；将 AIP 和相关的包描述转换成 DIP 集合并且将这些 DIP 存储在物理上分散的介质上,通过数据发布过程传递给用户，这一转化过程的复杂性因 OAIS 和用户的请求不同而异。

4.6 基于 OAIS 的数字记忆资源数据处理全过程

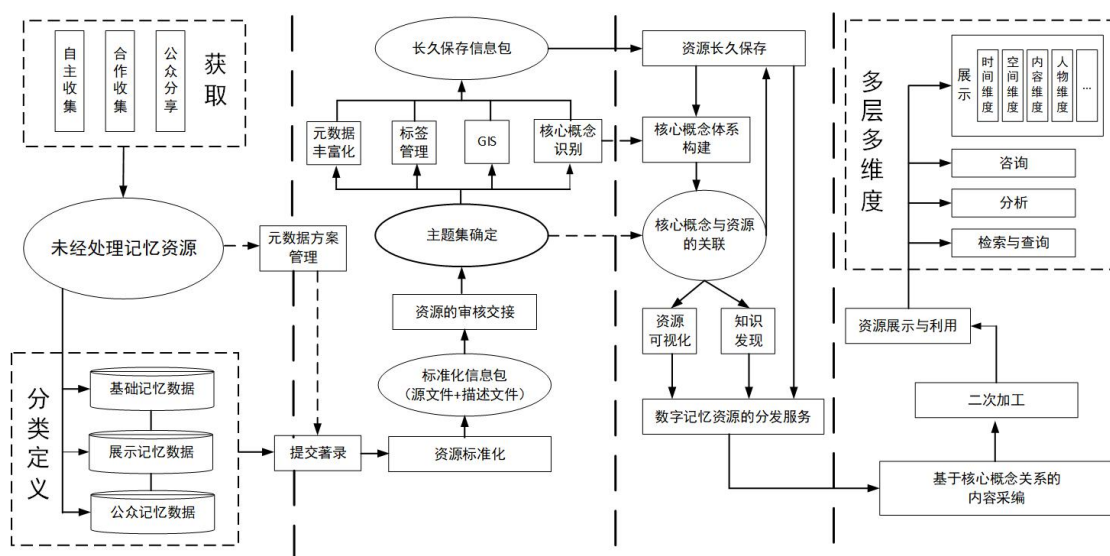


图 6 基于 OAIS 的数字记忆资源数据处理全过程

(1) 数据分类提交著录

将数据经过初步筛选清洗后，分类提交著录，强调与相应主题集的对接与责任主体问题，在此过程中尽量完善资源相关描述信息与元数据信息。

(2) 数据标准化与元数据丰富化

将已有的记忆资源数字化、标准化的过程，标准化的关键在于更具记忆

价值的资源的内容建立统一的著录和标引规范，统一元数据标准，并提供元数据标准的自定义设置。

（3）数据核心概念抽取及体系构建

核心概念识别，这一过程是对记忆资源内容要素的识别和提取，是建立“记忆工程”核心概念体系的重要前提。

（4）数据长久保存

基于核心概念的记忆资源关联可以揭示记忆资源之间的内在联系，并对关系进行语义化的描述。通过语义化的描述，可以揭示出记忆资源的潜在信息和价值。

（5）知识发现利用

基于核心概念体系可以实现资源的有效聚合，资源可以按照一定的类别实现多维度的汇集，如时间维度、空间维度、内容主题维度、人物维度等，进而实现资源的多维度展示，还可提供检索查询、分析、咨询等多层服务。

（6）数据新增/更新

考虑对基础记忆数据的本身的新增，以及高校记忆数据的外部补充，按照资源数据处理流程进行处理，焕新展示记忆数据，实现整个数据架构的生态发展。

5 面向数字照片的档案智能编研

5.1 档案编研的基本理论

5.1.1 基本概念

档案编研，是根据利用需要，依据一定的原则、方法和步骤，在研究档案内容的基础上，对档案文献进行收集、筛选、加工，形成不同形式的出版物，供各方面利用的一项专门工作。档案编研是历史档案开发的有效途径。

档案编研工作主要有两方面含义，一是“编”，即编纂；二是“研”，即研究。编是研的基础，研是编的深化。与专门的学术研究和史学研究不同，档案编研一方面侧重研究分析，整理、汇编和公布史料，配合编史修志等研究活动，主动提供利用服务；另一方面，对已经出版的历史著作进行史实印证，研究一定范围的历史问题，撰写历史著作，参加编史修志等活动。

5.1.2 原则与内容

档案编研工作在开展中需要遵循编研结合、依法合规、方便利用等原则，同时，编研工作内容要服从于编研工作的目的和意义，主要包括以下方面：

(1) 熟悉馆（室）藏档案资源构成，研究馆（室）藏档案资源内容

在开展编研工作时，首先应梳理馆（室）藏档案资源，包括馆藏的主要构成、载体形式、起止时间、特色馆藏等。只有对馆藏档案资源做到心中有数，才能实现高质量的档案编研。其次，对馆藏档案内容进行深入细致的研究。只有对档案内容进行细致研究，才能充分挖掘档案的价值与作用，编写出满足利用需求的编研成果。

(2) 形成不同层次的、形式多样的编研成果

档案编研工作的根本目的是通过不同形式的编研成果，挖掘和实现档案的价值和作用，更好地为利用者提供档案信息服务。

按照不同的分类标准，档案编研成果可以分为不同的类别：

按照加工层次划分，可以分为一次编研成果、二次编研成果、三次编研成果。一次编研成果是按照一定的要求，以档案文献原文为直接对象，经过收集、筛选、转录、校勘、编排、出版等环节，编制而成的反映档案原文的信息产品，比如各种文件汇编；二次编研成果是按照一定的要求，通过对档案原文内容进行浓缩、提炼、重新组合等加工，或用编者的语言表达，通过一定的体例编制而成的信息产品，如档案文摘；三次编研成果是在利用一次、二次编研成果的基础上，经过深入综合分析、研究编写出来的档案信息产品，如手册等。

按表达方式划分，可以分为文字式、图形式、数据式、声像式等。

按选题范围划分，可以分为综合型和专题型。

按体裁特征划分，可以分为汇编、文摘、索引、简介、史志等。

按载体划分，可以分为传统纸质载体编研成果，缩微品、电子出版物、网上出版物等新型载体编研成果，展品型编研成果，艺术型档案编研成果等。

5.1.3 基本流程

档案编研工作基本流程主要包括选题、选材、加工与编写、审核与出版等关键步骤。

步骤一：选题

（1）选题依据

选题就是选择档案编研的题目。题目是统领整个编研工作开展的主线，其他环节的工作都是依据所选题目来展开的。科学合理的选题应做到“三匹配”：

一是选题应以高校利用需求相匹配。首要满足高校内外部用户利用的现实需求，同时发现和研究档案利用规律，有预见性地选择符合时代发展要求的题目。

二是选题应以馆（室）藏档案资源相匹配。馆（室）藏档案资源是档案编研的基础，立足馆藏，结合自身特色和优势档案资源确立选题。

三是选题应以编研团队实力相匹配。不同的选题，编研工作难度各不相同。需要权衡考虑本单位的编研力量，选择与本团队实力相当的题目。

（2）拟写题目

编研选题不宜过长，用词准确精炼，具有针对性。

（3）拟制选题大纲

题目拟定后，根据选题方向与编研目的，拟制选题大纲，细化编研内容点。

（4）制定工作计划

以上所有工作完成后，根据编研力量，围绕此选题制定工作计划，明确人员、时间、负责内容安排。

步骤二：选材

在确定选题后，要通过各种途径查找档案，逐篇研究档案内容，考订文献真伪、形成时间，判断文献的价值等。选材确保材料真实可靠、材料齐全完整、材料详略得当与重点突出。

（1）选材的原则

档案选材应秉承全面查找、宁多勿漏、博约得当的原则，确保档案材料能够充分支撑档案编研。

（2）方法和步骤

首先，立足馆藏，对档案机构内档案文献进行查找；其次，放开眼界，对档案机构外档案文献进行查找。在这个过程中，把握材料线索“顺藤摸瓜”，并充分利用各种检索工具、数据库等全面获取资料。

步骤三：加工与编写

国家档案局
www.saac.gov.cn

材料选择后，就要对所选档案文献进行加工，根据最终形成的编研成果的不同，对档案文献加工的过程也有所不同，一般表现为对档案文献字句的勘正与恢复、划分段落与标点、修拟标题与正文的删节、对文献内容的提炼等。无论以何种形式对档案文献进行加工，都应坚持“维护档案文献的原貌，忠实文献内容的原意”，在真实、客观、准确的基础上进行。经过加工后的档案文献，须按照一定的体例进行系统化编排和格式化处理，形成便于利用者利用的编研成果。

此外，编研工作的程序还包括辅助性材料的编写等环节。不同编研成果的编写既应遵循编写过程中的统一要求和规律，又应满足各自的形式和特点，才能形成有价值、高质量的信息产品。

步骤四：审核与发布

编写形成档案信息产品后，需要对其进行审校与勘误，确保文字无误、格式正确、参考引证资料有源可溯，以纸质化或数字化的形式进行内容发布。

5.2 照片编研的记忆价值分析

数字照片是严格区别于纸质照片以数字化形式展现记录内容的一类照片。具体来说它是指用数字成像设备拍摄获得的，以及以扫描纸质照片及其胶片方式获得的，以数字形式存储于磁带、磁盘、光盘等载体，依赖计算机等数字设备阅读、处理，并可在通信网络上传送的静态图像文件。

作为信息的重要表现形式之一，数字照片因其各自具有的价值特性不同而归属于不同资源类别。随着“记忆工程”的广泛开展，加上照片数字化、数字照片保存技术的不断改进，数字照片内含的记忆价值日渐凸现。选取其中记忆价值突出但尚缺少相应价值保护的数字照片为研究对象（见下图阴影部分），分别从记录形式、内容、情感、视角四个维度对该部分数字照片的记忆价值进行分析，并归纳总结出以下四点。

国家档案局
www.saac.gov.cn

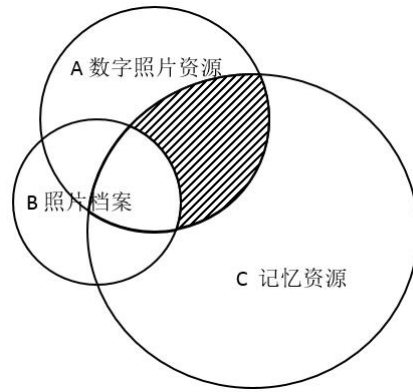


图 7 研究对象韦恩图

5.2.1 丰富记忆形式，描绘客观事实

数字照片通过显影技术将现实生活中的实时场景记录下来，存储于磁性载体材料并以数字化形式在电子设备上完整复刻呈现。这一特点弥补了人脑在记忆上容量有限、保真性难以保证、受寿命限制有保存时限以及在交流上受时空限制等固有缺陷。加之数字照片相对于传统文字记录的方式直观生动地再现了记忆原貌，提高了记忆内容的可读性，大大降低了失真度，减少了内容歧义，是记忆的有效载体。

5.2.2 补校恢复记忆，提供事实参考

历史典籍的编纂往往需要大量的史实材料辅以参考，一方面为写作、研究提供物质基础，另一方面实现多方资料相互考证，补全缺漏校对事实，还原记忆。数字照片大多反映某一时期的社会实践活动，经验证后可以作为补校与恢复记忆原貌的参考资料，并且数字照片的原始性一经确认，其内容真实性更胜于文字材料，对记忆的补校与恢复作用突出。

5.2.3 串联古今记忆，呼唤情感回归

数字照片的记忆价值不局限于解决眼下的现实问题，其价值的发挥更多在于激活记忆，引起情感共鸣。通常情况下，出于个人主观意识产生的数字照片往往记录动态事件中反映事件主题、最具代表性的一幕，常被作为一段记忆中的关键一环用以追溯往事串联记忆。特别是一旦当这一瞬间画面展现出来，它就能发挥以瞬时唤醒一件往事、一段记忆的功效并通过画面产生代入感引起观者情感共鸣。

5.2.4 多视角展现记忆风貌，体现公众参与地位

记忆视角下的数字照片打破传统官方独大、档案馆藏为主的记忆展现方式，

国家档案局
www.saac.gov.cn

更多地融合了公众记忆内容以公众视角讲述记忆历程较全面完整地反映了某一地区的集体记忆。摄影技术的普及使社会各领域的生产活动无一不暴露在公众的镜头之下，数字照片的反映内容贯通摄影发明伊始至今的城市巨大发展变迁时期，大到社会政治、军事、经济的巨大变革，小到胡同、邻里生活的家长里短，或独具代表性地记录客观事物某一时刻的形态，或完整连续记录社会组织及个人的实践活动共同组成了完整的社会印记。

5.3 智能化编研模式研究

5.3.1 数字档案编研的“记忆立方”

5.3.1.1 数字资源记忆立方

如下图所示，我们可以将数字资源进行立体式的“切分”，共分为三个维度：载体维度、来源维度、内容维度。

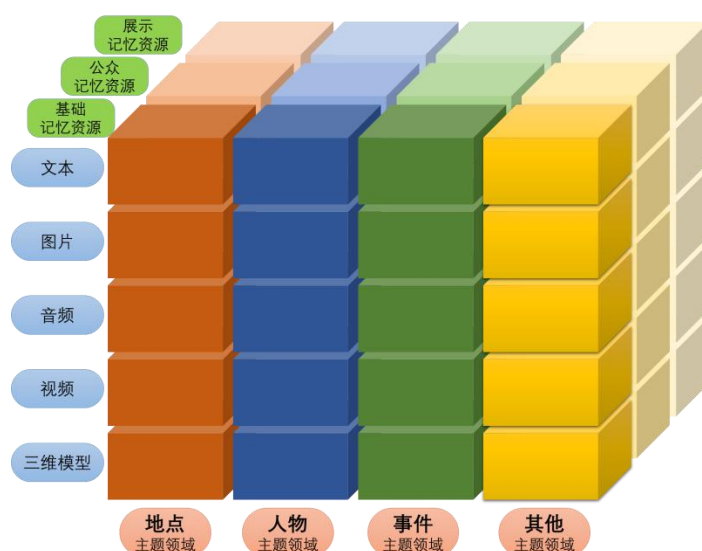


图 8 数字资源记忆立方

- 来源与用途维度：从数据的来源与用途可以分为基础记忆资源、公众记忆资源、展示记忆资源
- 载体维度：根据数字资源数据的载体形式不同，可以分为文本资源、图片资源、音频资源、视频资源、三维模型等。
- 内容纬度：按照记忆资源数据所描述内容的不同，可以按照内容维度进行分类，具体的分类内容需要根据实际情况进行判定，其分类内容类似于：地点主题维度、人物主题维度、事件主题维度等，也可以是

具体的内容主题。

5.3.1.2 面向专题的“切片”

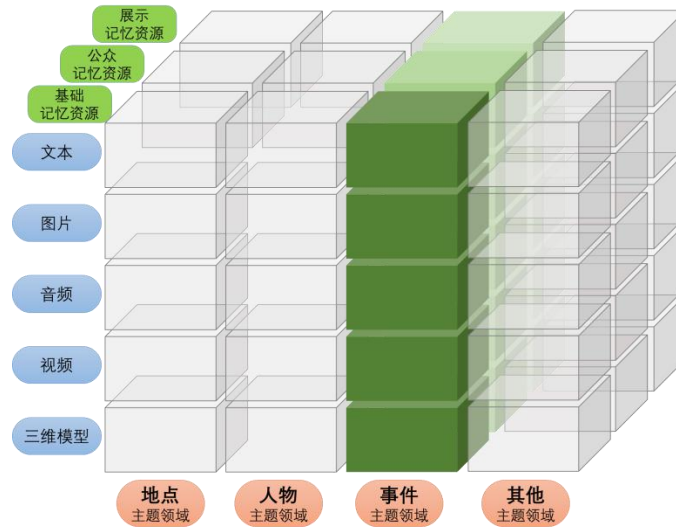


图 9 面向专题的“切片”

该种“切片”方式，是将数字资源按照主题内容进行划分，不论其来源或载体类型，统一进行组织加工，形成可展示的产品或服务。

5.3.1.3 面向载体的“切片”

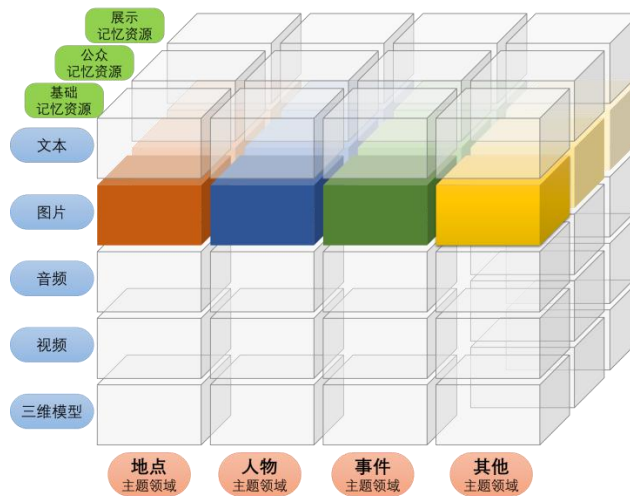


图 10 面向载体的“切片”

该种“切片”方式，是将数字资源按照载体维度进行划分，例如将所有照片资源整合出来，利用照片内容进行记忆的展示与传播。

国家档案局
www.saac.gov.cn

5.3.1.4 面向来源与用途的“切片”

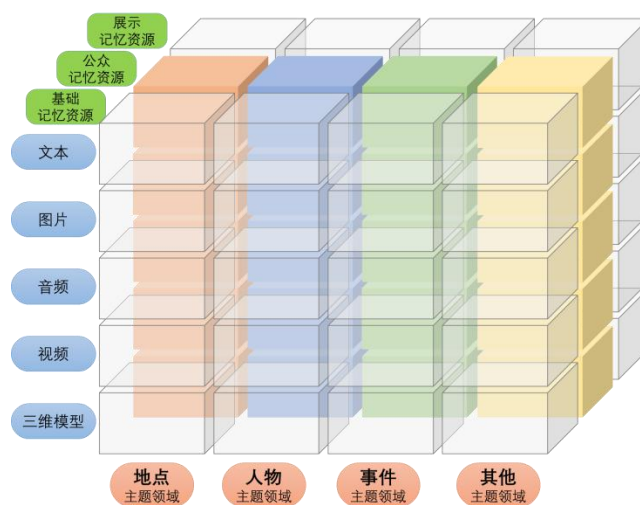


图 11 面向来源与用途的“切片”

该种“切片”方式，是将数字资源按照来源与用途维度进行划分，基础记忆资源可以作为综合查询系统的数据基础（也可将全部资源作为数据基础），共研究者查询基础性的记忆资源数据，开展相关的记忆研究；而公众记忆资源则可以单独作为一种聚类，专门展示公众个人的记忆资源，构建城市公民的个人记忆，并进行传播扩散；展示记忆资源是从其他两种记忆资源转化而来，将其统一聚类可以便于其他终端复用。

5.3.2 智能解析：基于记忆点的智能化资源关联

实现对数字资源的关联，需要构建规范的关联本体，使得系统资源都对应相应的记忆点，从而将资源库资源按照内容特征进行分类，构建实体之间的关联关系，再将资源与实体进行关联，实现资源之间的关联。这种关联模式实现资源的知识发现与深度价值挖掘，通过关系网络直观发现本体或资源间的相关情况，对资源的专题化聚焦和主题性组织价值巨大。通过数字化手段对大量资源内容或著录项进行识别、分析、关联、整合，实现智能化自动采编，大大提升后台管理系统的运行管理效率。

国家档案局
www.saac.gov.cn



图 12 记忆点关联资源示例



图 13 数字资源库本体对象关系网络示例

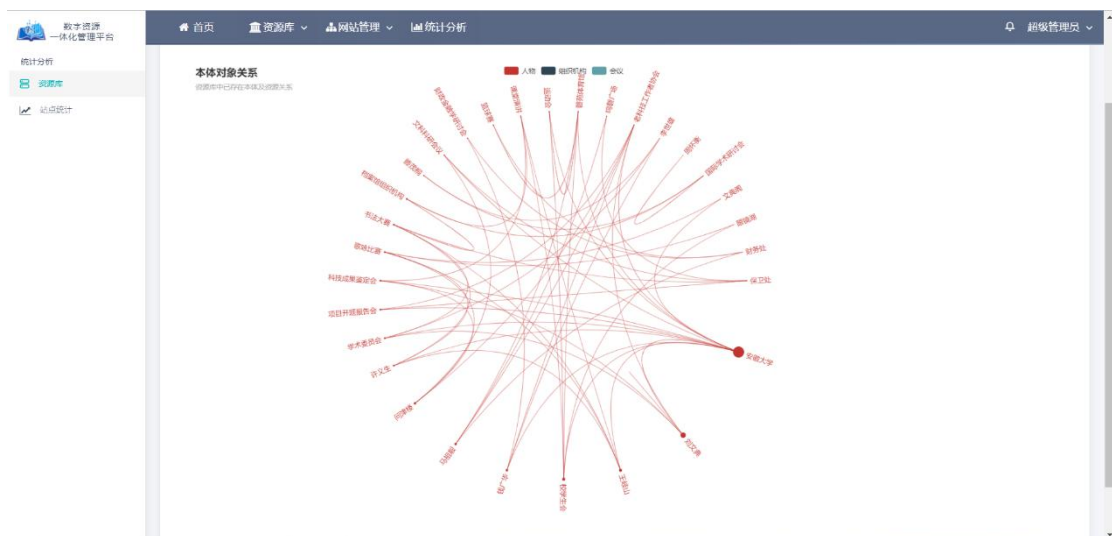


图 14 资源本体对象关系呈现示例

5.3.3 基于时空维度的智能化编研

基于时空维度的智能化编研是将照片资源按照照片的拍摄时间和拍摄地点，自动化映射到一张地图上，通过限定时间、空间的方式浏览制定照片资源。效果如下图所示。

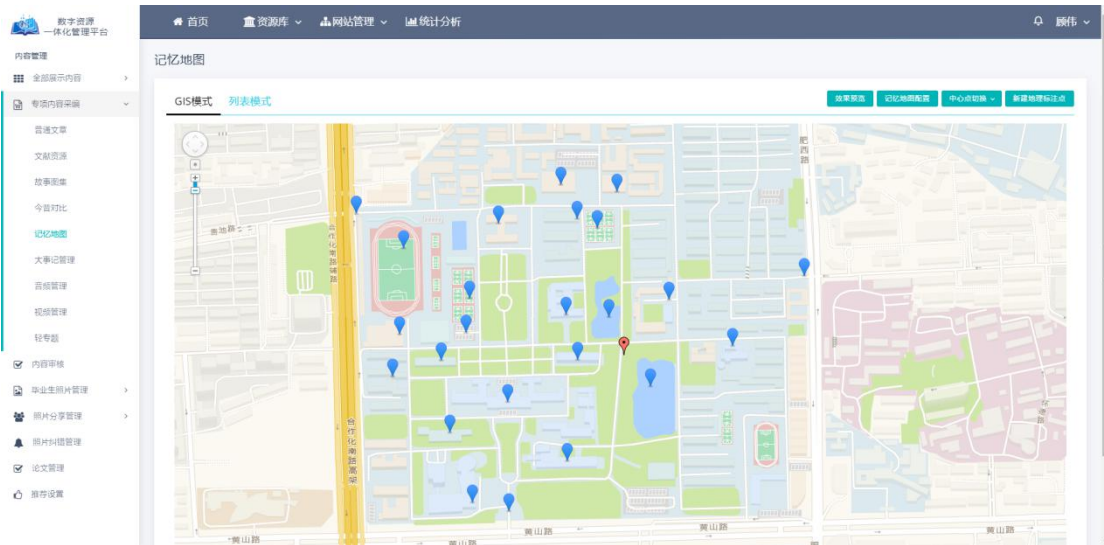


图 15 基于时空维度的编研界面

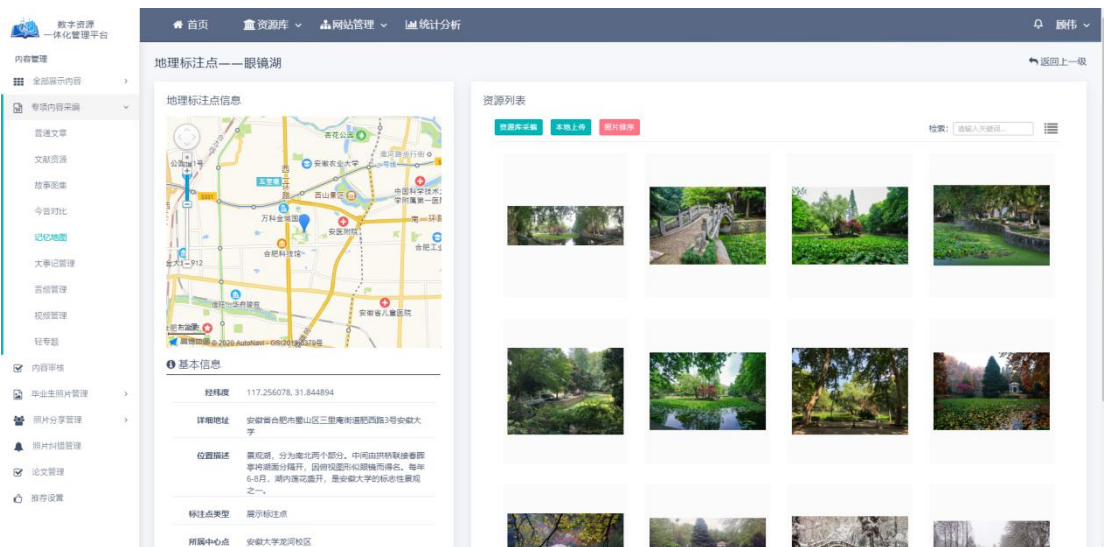


图 16 地理标注点管理界面

国家档案局
www.saac.gov.cn

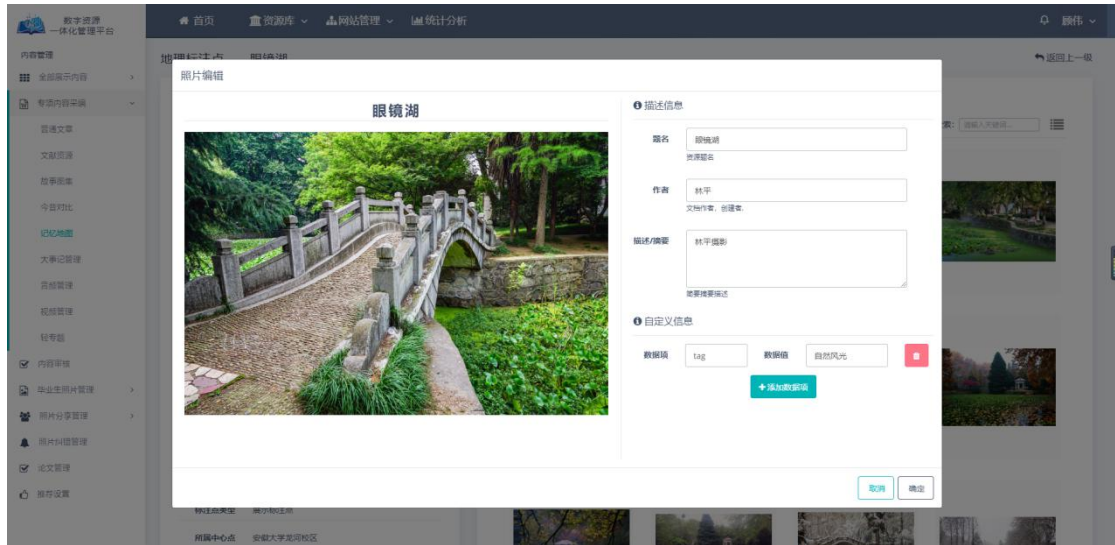


图 17 照片素材管理界面

如上图中的效果，地图上的每个点代表一个地理标注点，每个地理标注点下汇总了与该点相关的全部照片资源，包括经过二次采编的图集资源等。地图下方的时间轴可以拉动限制时间区间，并在地图上显示相应时间区间的资源点。点击具体的地理标注点即可访问该点一定时间范围内的照片资源。由于安徽大学具有多个校址，此处在实际设计开发过程中，会设计多个校区之间的切换功能，校区变化也将随时间一同变化。

5.3.4 基于主题维度的智能化编研

基于主题维度的智能化编研是基于对照片资源内容要素的关联性，进行某一主题内容的编研活动，如下图所示，是一个主题内容的照片故事编研。照片故事是通过一组照片，以图为主，以文字为辅的描述方式，进行照片资源的组织。



图 18 故事图集列表



图 19 故事图集编辑界面

5.3.5 基于场景维度的智能化编研

基于场景维度的智能化编研，是通过照片场景的对比，从海量照片中找到具有相同地点场景的古今照片，进行同一个场景下的照片对比性组织。例如，主教学楼十年前的照片和现在的照片，同一个角度，同一个建筑，对比产生强烈的历史变迁感。该模式下，可以将多个场景、多组对比照片进行组织，形成一整套今昔对比编研成果。

国家档案局
www.saac.gov.cn

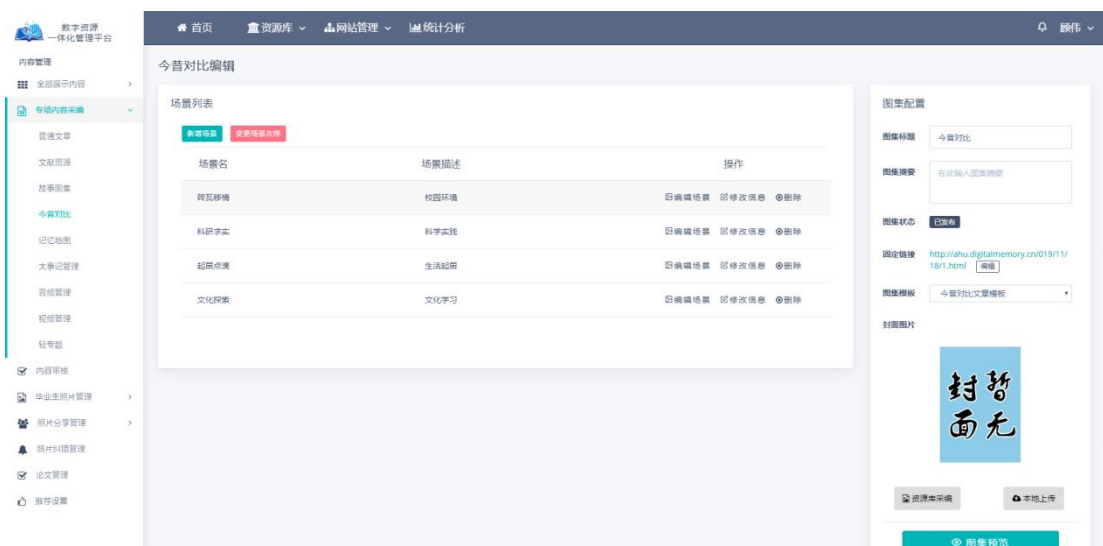


图 20 场景列表

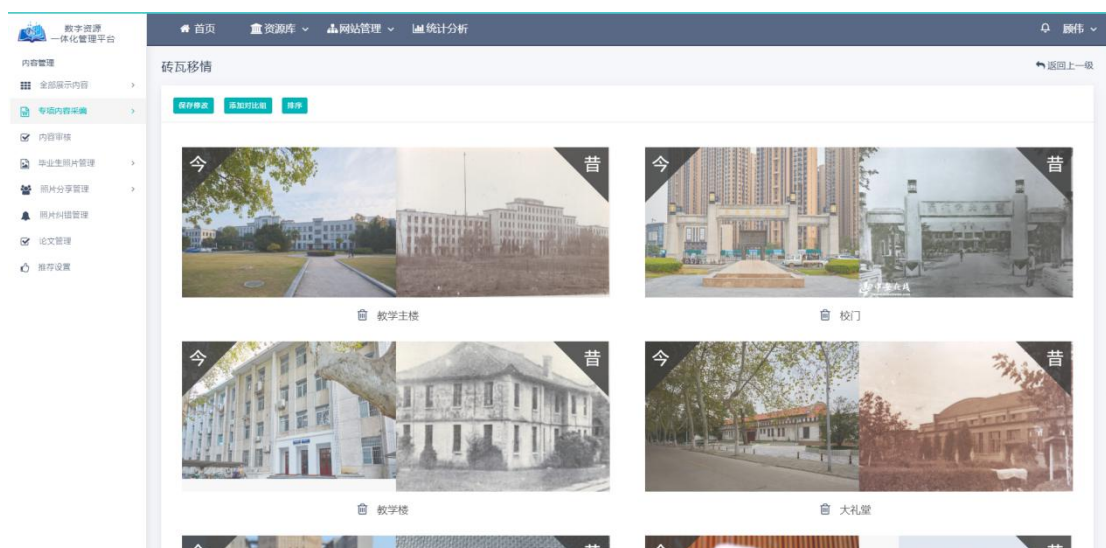


图 21 对比组管理

5.3.6 照片档案编研可视化设计方案

安徽大学档案馆通过校史编研、多渠道征集等多种渠道管理照片档案资源，提供了包括安徽大学人才培养、科学研究、学科建设、师资队伍、国际交流与合作、党建等多类档案资料；另外，校史馆的展览资料也被重新扫描，呈现了包括亲切关怀、校史沿革、今日安大、校友与母校、校友风采、馨苑校区建设和徽文化专题七个方面的内容。

全面的数字资源载体内容，在“安徽大学数字记忆网”创新呈现出基于时空维度的瀑布流、时间轴和基于主题维度的故事图集、今昔对比、毕业生照片墙、名人专题等特色展示方式。



图 22 网站首页

(1) 基于时空维度

网站通过 GIS 创建记忆地图，在地理标注点上呈现地标建筑图片。对于具体某一地标建筑的大量图片资源，通过瀑布流和时间轴两种展示方式呈现。瀑布流将照片档案资源平铺展示，时间轴则按时间顺序展示图片，体现历史长河的厚重感。

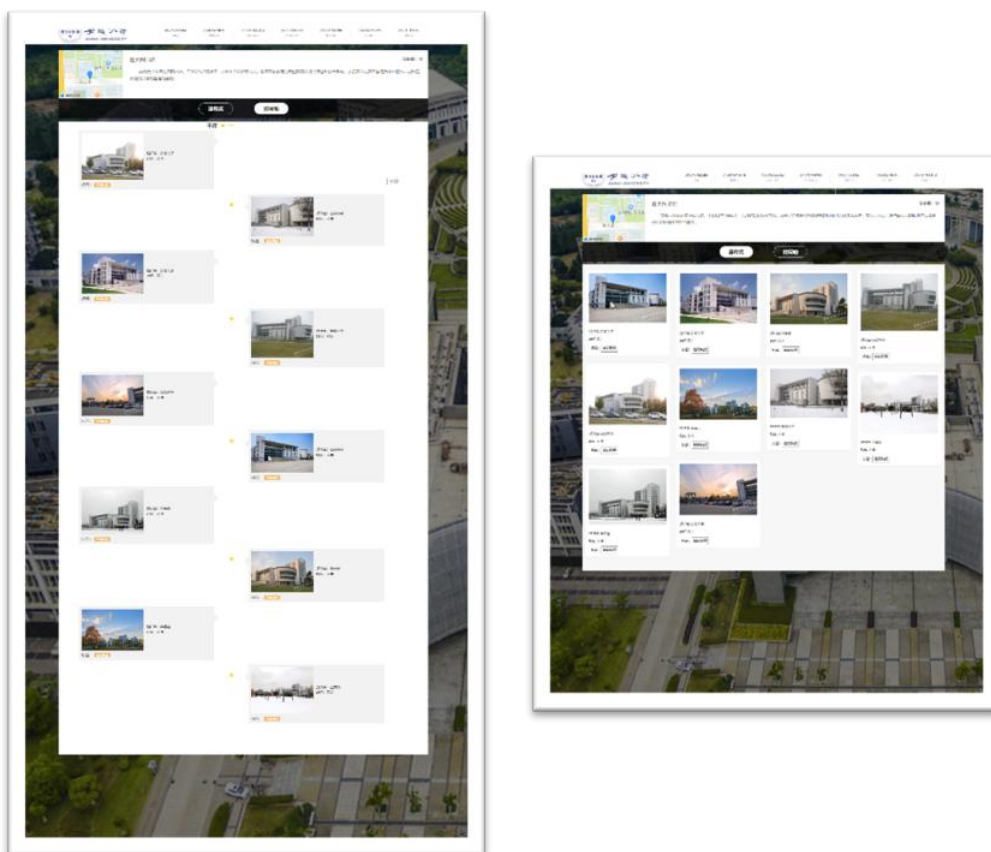


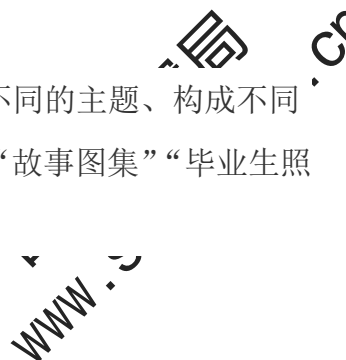
图 23 时光轴与瀑布流



图 24 GIS 地图

(2) 基于主题维度

照片档案资源的元数据内容复杂，每一张图片对应不同的主题、构成不同的逻辑故事线。因而基于内容维度按主题组织图片，有“故事图集”“毕业生照片墙”“名人专题”等方式。



“故事图集”是网站照片档案资源展示的主要形式之一，故事图集将一组照片档案资源汇集，通过文字的描述和照片顺序的排布，使得一组照片能够描述一个完整的主题或者故事。



图 25 故事图集

“毕业生照片墙”是专门为毕业生照片提供的资源展示方式，可以通过查询、筛选的方式，找到指定年份、专业的毕业照，使得校友们可以快速找到当年的“我”和“我们”。

“名人专题”则是为安徽大学知名学者提供的专题式资源展示栏目，通过故事图集等方式，描绘各位学者的辉煌成就和历史印记。



图 26 毕业生照片墙



图 27 名人专题

(3) 基于场景维度

同一个场景下的不同年份照片可以为人们带来强烈的历史变迁感，通过展示同一个场景下的古今照片，使用户有一种穿越古今感受历史变迁的感觉。在展示形式上，要突出对比角度的一致性、对比地点（场景）的变化性、对比观感的反差性等。

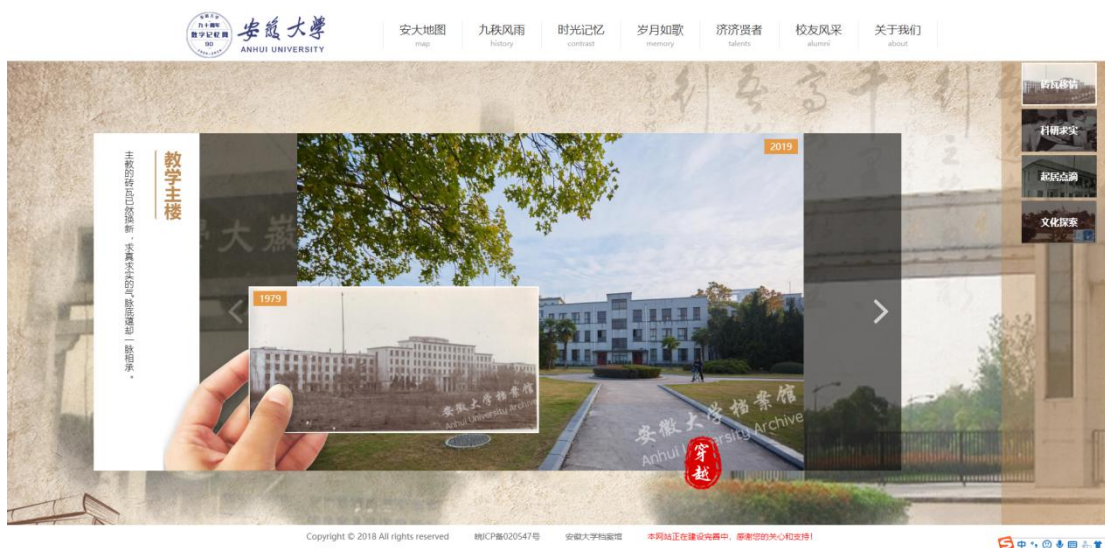


图 28 今昔对比

6 高校历史照片档案智能编研系统设计

6.1 平台控制台设计

控制台呈现系统资源量、文章量的统计和总览，以及对前端网站站点的访问监控，包括一些经过处理的可视化图表和详细的资源信息。控制台具体包括

四个部分，站点概览——对站点资源和访问情况的集中展示；资源提交情况——记录不同类别的资源提交趋势和平台用户提交资源的贡献操作；资源库管理情况——直观呈现资源库中资源的状态比例和所采编文章的情况；网站管理情况——呈现网站前端的近期访问情况、趋势统计和热门访问页面情况。

6.2 资源库管理功能设计

6.2.1 提交模块功能设计

资源提交模块包括元数据方案设置、提交资源情况分类概览和具体的资源提交界面。供用户提前设置资源分类、资源标签，并进行资源的提交著录操作。

6.2.1.1 资源提交功能设计

该板块主页面提供按载体分类提交和打包提交两种方式。

在照片档案项目中，资源提交仅涉及按载体分类提交资源中的照片模块，用户可点击①进入图片资源列表，点击左上角“资源提交”进入资源提交页面。

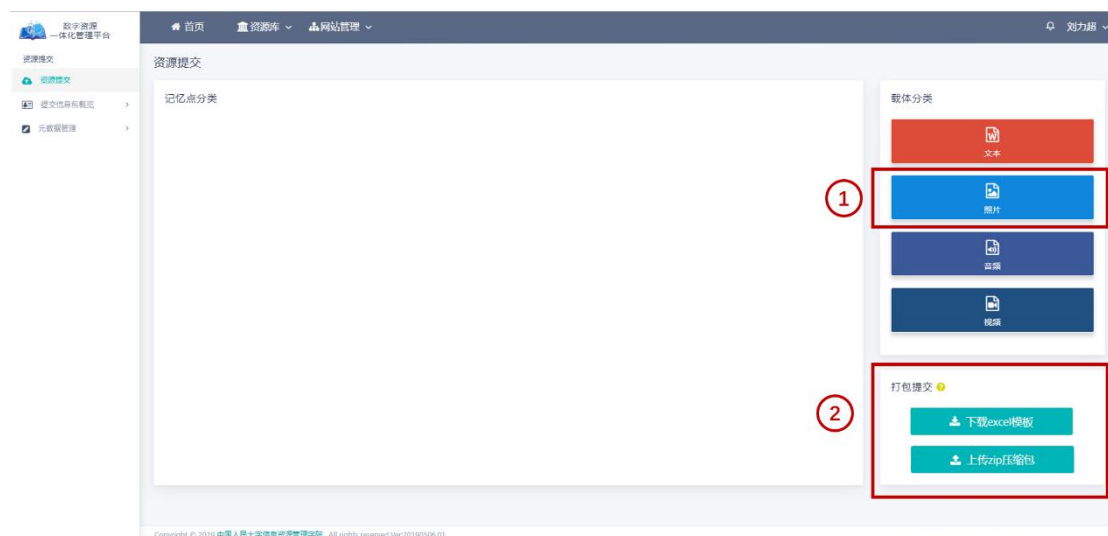


图 29 资源提交模块主页面

国家档案局
WWW.SAAC.GOV.CN



图 30 图片资源列表



图 31 资源统一提交入口



图 32 待录入资源列表返回编辑页面

图片资源编辑页面左侧显示图片，右侧供用户进行元数据著录，包括内容描述类、外部属性类、版权信息描述类和自定义描述类元数据信息，确认后保存变更即可完成单个图片的资源提交。

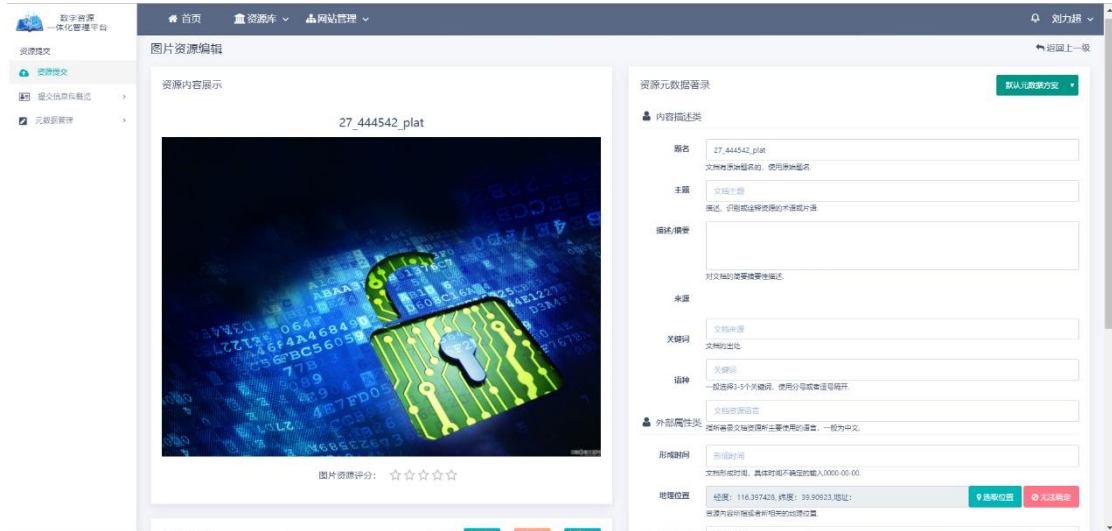


图 33 图片资源编辑页面

另外，系统可实现批量提交的三种模式。

(1) “数字资源组+元数据著录表”的打包提交方式

在资源提交界面②处打包提交板块，用户可点击“下载 excel 模板”选择元数据方案，预先下载某一元数据方案的著录模板（元数据著录表），将使用相同元数据方案的一组数字资源提前填写到元数据著录表中，将数字资源与元数据著录表一同打包成一个压缩包，一次性提交到系统中，系统将自动识别元数据著录表中的元数据信息，并一一对应的将元数据著录到对应的数字资源信息包中，免去逐份数字资源著录所带来的大量工作。