档案文化智慧数据资源建设——河南省档案馆馆藏中福公司档案整理开发研究之二*

郝伟斌，王君仪，段燕鸽

摘　要：中福公司作为中外合资企业，在中国经营期间留下了极为珍贵的档案资料。中福公司档案作为价值丰富的历史文化档案资源，国内外学者已对其开展不同层次与角度的研究，获得了丰富的研究成果。国家文化大数据体系建设工作的开展与档案数字化工作逐渐兴起，基于此，以河南省档案馆主导的中福公司档案整理与开发项目为契机，梳理分析智慧数据为中福公司档案智慧数据资源建设带来的契机以及中福公司档案智慧数据资源建设可行性，并在此基础上以工程化与系统化的模式框架为基础，构思包含数据获取、数据转化、数据关联与数据应用的中福公司档案智慧数据资源建设技术路径。

关键词：中福公司；智慧数据；档案文化资源；资源建设

Abstract: As a Sino-foreign joint venture, Zhongfu Company has left a very valuable Archives during its operation in China. As a valuable historical and cultural Archives resource, scholars at home and abroad have carried out different levels and angles of research, and obtained rich research results. Based on this, the development of the national cultural big data system construction and the digitalization of Archives are gradually emerging, taking the Archives arranging and development project of Zhongfu Company, which is led by henan Province Archives, as an opportunity to sort out and analyze the intelligent data for the construction of the Archives intelligent data resources of Zhongfu Company and the feasibility of the construction of the Archives of Zhongfu Company, and on this basis, based on the model framework of engineering and systematization, the concept includes data acquisition, data transformation, Data association and data application of Zhongfu company file intelligent data resource construction technology path.

Keywords: Zhongfu company; Smart data; Archival cultural resources; Resource construction

DOI:10.15950/j.cnki.1005-9458.2022.01.016

智慧数据作为数据科学领域的新概念，通过挖掘用户需求，依托细粒度的知识组织与表示，以语义化及可视化为进阶，拓展数据知识化应用，形成数据的智慧化高阶价值呈现形态。随着中福公司档案整理与开发项目的逐步推进，档案中蕴含的潜在价值逐渐得以开发与显露，有必要针对中福公司档案已有的数字化基础融合智慧数据理念与技术开展进一步探究，通过本文研究，将理论与实践相结合，以中福公司档案智慧数据资源建设为例，以期助力于档案文化资源建设理念创新，对于档案文化智慧数据资源建设工作起到进一步推动作用。

1 智慧数据处理是档案文化资源建设的新手段

1.1 激发档案文化隐性价值。档案智慧数据作为对档案实体深入挖掘得出的高阶价值呈现，在档案文化资源建设工作中引入智慧数据技术与理念，智慧数据自身所具备的价值增值性使得档案资源隐性价值——档案数据价值得以充分开发。

1.2 细化档案文化数据粒度。档案粒度是指在不同角度与层次对档案资源细化后产生的数据元素基本构成单元。智慧数据所应用的知识发现技术，可通过数据挖掘、机器学习、深度学习等方式展开自动分析，快速洞察细粒化数据的隐藏关系，对数据进行预处理，实现知识单元离散化、细粒度知识组织与揭示服务的精准化、语义关系丰富化等。将智慧数据相关技术理念融入档案文化资源建设，获取合适的档案数据分化理念，推进档案数据粒度细化，形成档案数据结构中相对独立的、具有完备知识表达的、最细粒度化的概念模型，提高档案数据知识主体构建工作的效率与精确率，随之通过语义丰富化，实现数字资源间语义关系的建立和扩展，促进大规模档案资源之间的关联融合，提高档案文化资源的可用性和共享性。

1.3 加速档案文化智慧发展。在技术方面，智慧数据技术作为数字化、数据化、智慧化等阶段关键技术的融合，包含数据管理技术、数据安全技术、语义化技术、可视化技术等，可以促进档案行业充分利用数据挖掘、分析、关联等适用性新型技术、智能化设备与数字化平台，丰富档案文化资源自身语义，实现数字档案之间语义关系的建立，推进档案资源向数字化—数据化—智能化转变。在思维方面，智慧数据所包含的数据意识与态度、数据处理思维、智能平台化思维、价值取向等数据素养，能够促进档案工作思维泛化，推动档案文化资源精细化建设，不断激发档案文化资源之中的

巨大价值，加速档案文化领域的智慧化发展。

2 中福公司档案智慧数据资源建设可行性分析

2.1 自身价值。中福公司是西方列强于中国近代时期在华投资创办的一家大型外资企业，在中国经历了福公司独资经营、福中总公司合营、中福两公司联合办事处三个阶段，主营煤矿，兼营铁矿、铁路、桐油、特种矿产品等业务，其活动范围涉及北京、天津、山西、河南、湖北、湖南等地区。中福公司档案作为英国福公司在中国从事政治、经济、教育等活动直接形成的具有保存价值的历史记录，分散保存在河南省档案馆、湖北省档案馆、重庆市档案馆等地区，具有内容丰富、载体多样、类型丰富、资源地位显著、史料内容充足、研究价值独特等特点。[1]其作为一座档案的“富矿”，能够为寻求史实、开展学术研究提供一手史料，为推进社会主义爱国教育提供基本材料门径。[2]

针对中福公司档案进行智慧数据资源建设，以用户动态化、多元化、及时性的信息需求为中心，充分运用数据技术与智能技术，打造中福公司档案智慧数据资源知识库，是实现中福公司档案文化价值、学术价值与教育价值最大化体现的重要途径。

2.2 基础优势。河南省档案馆对中福公司档案进行整理与数字化开发已取得较为显著的成果，为中福公司智慧档案数据资源建设的开展打下一定的基础。一方面，数字化处理与加工，依托中福公司档案实体形成了较为完备的数字化资源。另一方面，河南省数字档案馆建设，为中福公司档案智慧数据资源建设提供了智能化技术支撑与数字化环境优势。

2.3 理论支撑。2018年，中国人民大学钱毅教授首次提出“三态两化”理论。“三态”指的档案对象管理空间的模拟态、数字态与数据态，模拟态注重维持实体有序与存贮空间安全，数字态注重保证数字态对象可读性，数据态注重维护数据态对象的可理解性。[3]钱毅教授强调，以维护语义完整为主的档案数据态保存则成为亟须关注的重点问题。

中福公司档案现如今已通过派生方式实现存量档案数字化、完成数字共享平台建设。同时，编纂了《中福公司档案史料汇编》，拍摄了《他们特别能战斗》文献纪录片。基于信息化深入发展、数据驱动普遍出现、档案管理对象维度收缩、三态并存等社会发展情形，具备了依据自身深度开发利用的条件和需求，推进中福公司档案智慧数据资源建设的条件。

3 中福公司档案智慧数据资源建设模式

3.1 工程化项目驱动。“建”的目的在于“用”。建立档案文化智慧数据资源的工程化建设模式，采取过程性、流程化管理策略，可巩固阶段性建设成果，稳步推进建设项目的实施。

针对中福公司档案史料汇编项目的智慧数据资源建设工作，融合项目工程化思想，应明确项目整体目标，定位资源建设需求。以项目需求为导向，把控建设节点以确保建设目标达成的准确性和资源建设的完整性。

3.2 系统化多方协同。（1）主体引领。一方面中福公司档案智慧数据资源建设工作主体——河南省档案馆根据中福公司档案特点，研究确定中福公司档案资源建设方案，细化建设理念与工作节点；另一方面河南省档案馆积极推动多方协同，如档案修复与数字化协同、翻译与数据转化协同、数据关联发布与数据应用协同等，高效做好中福公司档案智慧数据资源建设工作。

（2）多方协同。中福公司档案智慧数据资源建设全过程不仅包括针对国内外中福公司档案史料进行调研、收集的档案资源准备工作，还包括馆藏档案的修复、分类和翻译，以及中福公司档案数字化加工和资源平台建设等。河南省档案馆仅依靠自身力量难以高质量完成，需要研究、翻译、修复、数字化、平台搭建等多方技术团队协同完成。[4]

从系统论的角度来看，中福公司档案智慧数据资源建设以技术参与方为依托，对传统档案资源进行处理，将其以一定的层次与结构有机结合起来，作为该生态体系的“骨骼”，并以文化为题，赋予其独特内涵，作为该生态体系的“血液”，使中福公司档案资源在该体系中得以循环流动，共同构成系统协同性智慧数据资源建设模式。

4 中福公司档案智慧数据资源建设技术路径

智慧数据是信息资源的高级组织形态与表达方式，数据的结构化、语义化和关联化程度相比现有信息资源组织程度更高，是数据科学理论体系中的新概念和信息资源建设的新方向。技术路径包括数据获取、数据转化、数据关联和数据应用四个方面。数据获取方面，重点在于结构化转换，构建中福公司档案资源数据库；数据转化方面，通过五大概念模型细化中福公司档案资源类别，形成细粒度的档案知识元以构建档案知识本体；数据关联方面，依照中福公司档案主题词表，利用语义组织技术实现数据资源的深度标识；数据应用方面，实现主题检索、知识推荐与智慧服务，以个性化、多样化的形式呈现中福公司历史图景。

4.1 数据获取。构建资源数据库时所处理的资源对象可以大致分为非结构化资源、半结构化资源和结构化资源三种。构建中福公司档案资源数据库首先针对数字化后的中福公司档案图像进行OCR识别，结构化档案资源，也就是分离档案资源图像层与文本层，增加其结构与内容的分离程度。文本主要由内容、结构组成，内容表述信息的语义含义，是文本的核心部分，也是获取语义信息的重要来源；结构则用以支持语义的内容表述，从句法结构中有效地识别词语，并建立文本概念之间的对应关系是获取档案资源语义信息的关键途径。

4.2 数据转化。“本体”一词源于哲学领域，且长期以来存在着许多不同的用法。在计算机科学领域，其核心意思是指一种模型，用于描述抽象概念、概念的属性及其之间的各种关系。档案内容的语义集中体现在时间、空间、人物、组织和事件五大方面，借由“本体”的内涵引申至档案实体，构建各实体所对应的抽象概念模型，针对中福公司档案资源中包含的这五大数据属性形成档案知识本体。

在档案文化资源本体构建过程中，人们难以做到对实体别名的穷举式构建抽象概念模型，别名与目标对象之间缺少显式的链接关系，实体名称的变更将会导致档案链之间的断裂，最终造成档案文化资源抽取时无法保证抽取结果的查全率。[5]

中福公司档案时间跨度较大、语言体系混杂，至于中福公司档案资源的本体构建则需要解决概念模型的统一表述问题。

在时间方面，中福公司作为中外合资企业，业务活动中形成档案所采取的纪年方式也存在差异，确立中福公司档案本体的时间概念模型时应采取统一纪年形式，确保时间描述准确；在空间方面，中福公司从注册成立到终止经营历经半个世纪。其间，存在历史环境变化导致的地名变更，档案中所记载的关于地理位置的内容或存在“一地多名”现象，为此需做到根据可考历史事实，梳理地理位置名称的演变情况，采用统一的空间位置描述语言，保证空间概念模型的准确性；在人物方面，中福公司档案存在同一人物拥有多种不同称谓的别名现象，如孙越崎与其原名毓麒，因此在构建人物概念模型时需建立人物实体别名间的关系，明确人物概念模型的称谓指代，确保其唯一性；在组织方面，与人物概念模型类似，中福公司档案中出现的社会组织，需要按照一定的叙词表标准为其建立分类体系，建立每个社会组织不同名称代指的唯一标引符，确保每一社会组织名称代指在叙词表中存在与之相对应的标引符；在事件方面，针对中福公司档案实体所反映的真实事件详情构建事件概念模型。

以概念模型形式对中福公司档案资源进行分解，重组为细粒度的档案知识元从而构建档案知识本体，为后续中福公司档案资源的数据关联、发布与应用建立数据基础。

4.3 数据关联。借助语义网技术的档案文化数据资源语义组织是构建知识本体之间语义关系的重要环节，同样也是智慧数据资源建设区别于传统档案资源建设所在。

通过对中福公司档案本体进行语义组织，建立档案知识本体之间的逻辑关系，将数据与实物、数据与数据等关联起来，构建中福公司档案数据资源内关联，以关联数据的方式进行发布，从而形成一张巨大的档案资源语义数据网络。

在由河南省档案馆主导的中福公司档案整理与开发项目开展过程中，制定了《中福公司档案著录细则》与元数据方案，并建立了高频的人名、地名、货物名等英文与中文翻译对照表，参考ISO25964-2所制订的叙词表与其他词表映射的标准可形成中福公司档案主题词表。根据中福公司档案主题词表所设定的描述规则，针对中福公司档案知识本体概念模型中时间、空间、人物、组织和事件或者实物语义之间进行的标引，实现对中福公司档案内容、形式和管理特征的规范化标引，建立起中福公司档案在某一主题下档案汇集内不同文件之间的关联，形成关于这一主题的完整、详细的内容网络。

4.4 数据应用。专题检索。中福公司档案史料作为相关领域专家学者们长久以来关注的重点对象，中福公司档案智慧数据资源建设成果将为其带来学术研究的便利。利用经过语义组织与关联发布的中福公司档案智慧数据进行专题化档案知识检索服务，专注于用户的实际需求，按照不同的专题遴选有较强利用价值或潜在利用价值的档案智慧数据，最大程度满足用户专题档案知识的需要，档案利用者在进行检索时仅需输入所需档案关键词，即可得到该主题相关的查全率较高的档案数据以及其知识化组织成果，大大降低了专家学者们的检索成本，从而进一步促进中福公司档案的利用。

知识推荐。档案知识推荐服务作为档案知识化服务的一大组成部分，可以依据用户画像主动地提供有针对性的推荐服务，在档案服务过程中能够起到优化服务提供方式和改进现有服务手段的双重作用。用户以一定目的性查阅中福公司档案时，可为之推荐与其查阅目标相关的人物、地点或事件，引导和满足用户的知识需求，满足档案知识服务与用户需求的双向匹配，创新中福公司档案智慧数据开发与服务方式。

智慧服务。综合数字人文理念与知识图谱的技术方法，建设基于GIS技术的中福公司档案资源和文化共享知识库，根据不同主题将事物的空间数据和属性数据结合起来提供给用户，将该技术应用于中福公司档案文化数据的呈现，除提供必要的检索功能外，还可根据不同主题将事物的空间数据和属性数据结合，借助可视化技术展示，构建灵活的、开放的、延续的，集时间、空间、人物、组织、事件多项档案数据展示于一体的中福公司档案智慧数据资源的共享平台，形成完整的数据地图，将中福公司历史真实图景铺展于用户面前，尝试构建中福公司历史模拟环境，利用智能问答技术，增强人机交互体验，为其提供中福公司档案智慧化服务。

*本文系河南省科技攻关项目“数字人文视域下中福公司档案开发利用研究”（202102310307）阶段性成果。

参考文献：

[1]李宗富,崔白璐.国内中福公司档案研究回顾与展望[J].档案管理,2020(03):92-95.

[2]衡芳珍.英商福公司研究述评[J].河南理工大学学报(社会科学版),2010,11(02):242-249.

[3]钱毅.基于三态视角重新审视档案信息化建设[J].浙江档案,2019(11):18-21.

[4]李宝玲,朱兰兰.重构历史真实图景:河南省档案馆馆藏中福公司档案整理开发研究之一[J].档案管理,2021(03):7-10+14.

[5]夏天,钱毅.面向知识服务的档案数据语义化重组[J].档案学研究,2021(02):36-44.

（作者单位：郑州航空工业管理学院 来稿日期：2021-10-19）