|  |  |
| --- | --- |
| ICS  |  |
| CCS  |  |

中华人民共和国档案行业标准

DA/T XXXXX—XXXX

DA

档案专题数据库建设工作规范

Work Specification for the Construction of Archival Thematic Database

（征求意见稿）

XXXX - XX - XX发布

XXXX - XX - XX实施

国家档案局  发布

目次

前言 III

引言 IV

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 总体要求 2

5 组织管理 2

5.1 组织与人员 2

5.2 建设方式 2

5.3 实施模式 2

5.4 流程管理 3

6 技术保障 3

6.1 设备保障 3

6.2 系统功能 3

6.3 技术应用 3

7 选题与建设方案 3

7.1 选题提出 3

7.2 选题分析 3

7.3 确定选题 3

7.4 建设方案 3

8 数据采集 4

8.1 采集方案 4

8.2 采集方式 4

8.3 采集要求 4

8.4 数据归集 4

9 鉴定与审核 4

9.1 鉴定 4

9.2 审核 4

10 整理与著录 4

10.1 整理 4

10.2 著录 5

11 质量管理与验收 6

11.1 数据清洗 6

11.2 数据质量检查 6

11.3 验收 6

12 查阅利用与宣传推广 6

12.1 查阅 6

12.2 公开发布 6

12.3 查阅利用平台 6

12.4 产品开发 7

12.5 宣传推广 7

13 管理与维护 7

13.1 反馈评价 7

13.2 动态管理 7

13.3 维护备份 7

附录A （资料性）扩展著录（示例） 8

附录B （资料性）来源信息著录（示例） 9

附录C （资料性）使用限制信息著录（示例） 10

附录D （资料性）通用项目对应关系表（示例） 11

参考文献 12

1. 前言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由国家档案局提出并归口。

本文件起草单位：国家档案局科技信息化司、云南省档案馆。

本文件主要起草人：

1. 引言

档案专题数据库建设是专项工作、重大活动、突发事件档案收集、管理的重要途径，是高效组织、管理和利用档案资源的手段，还是深度挖掘和开发档案资源的方法。本文件立足档案专题数据库建设的研究和实践，提出档案专题数据库建设的组织管理和实施等要求，以规范档案专题数据库建设工作。有关数据库、信息系统开发的具体技术运用不纳入本文件范围。

档案专题数据库建设工作规范

* 1. 范围

本文件确立了档案专题数据库建设工作的总体要求、组织管理、技术保障、选题与建设方案、数据采集、鉴定与审核、整理与著录、质量管理与验收、查阅利用与宣传推广，以及管理与维护等。

本文件适用于档案主管部门、综合档案馆及机关、团体、企业事业单位档案馆（室）开展各类档案专题数据库建设工作。

* 1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 39784—2021 电子档案管理系统通用功能要求

DA/T 18 档案著录规则

DA/T 31 纸质档案数字化规范

DA/T 43 缩微胶片数字化技术规范

DA/T 46 文书类电子文件元数据方案

DA/T 54 照片类电子档案元数据方案

DA/T 62 录音录像档案数字化规范

DA/T 63 录音录像类电子档案元数据方案

DA/T 89 实物档案数字化规范

DA/T 99 档案数字资源备份实施规范

* 1. 术语和定义

下列术语和定义适用于本文件。

档案专题数据库 archival thematic database

围绕特定的对象、范围、内容等主题范畴，按照概念结构组织的档案及其相关信息的数据的汇集，它描述这些数据的特征及与数据对应的实体间的关系并支持一个或多个应用领域。

注1：GB/T5271.1-2000,定义01.08.05 改写。

注2：档案专题数据库的数据包含数据集、文本、图片、图形、音频、视频等结构化、半结构化及非结构化数据。

数据采集 data acquisition

根据档案专题数据库建设需要，按照一定规范，获取各种来源、形态的档案及其相关信息的过程。

数据清洗 data cleaning

发现在数据采集、传输、处理或存储过程中产生的存在错误、不完整、不准确、不一致、重复或含有噪声等情况的数据，并进行纠正的过程。

* 1. 总体要求

4.1 目标导向：应以服务党和国家工作大局，弘扬时代精神和优秀传统文化，记录重大活动、突发事件，服务民生，服务本行业、本单位工作为引领，按需建设。

4.2 统筹推进：应统筹目标任务、档案资源及保障条件等，制定档案专题数据库建设方案，按项目组织建设工作，量力而行，分步实施，可跨区域、跨层级、跨部门协同推进。

4.3 科学规范：应制定工作规范，技术方案，明确档案专题数据库建设流程、实施方法、数据规范和质量要求等。

4.4 安全合规：档案专题数据库建设全过程应当符合国家网络安全、数据安全、个人信息保护、保密管理、密码管理等相关法律法规要求；依法保护国家和社会利益，保护知识产权及合法权益。

4.5 可信可用：数据来源可靠，信息准确，内容完整，对国家和社会具有保存利用价值；数据组织管理有序，可读可用。

* 1. 组织管理
		1. 组织与人员

应明确档案专题数据库建设的管理、实施部门及人员。

承担管理职责的部门负责选题、制定工作方案和业务规范，统筹资金、技术等保障条件，负责监督指导、组织实施、验收评价等工作，可根据需要聘请有关专家和顾问。

承担工作实施职责的部门负责按照总体安排或合同、协议，开展数据采集处理、存储、利用等工作，负责系统建设与运维。应根据所承担的工作内容，合理配备档案收集整理、保管、开发利用、数据治理、信息系统管理等方面的专业人员。

* + 1. 建设方式

应根据工作实际，采取以下一种或多种建设方式：

a）档案汇集。围绕特定主题，筛选馆（室）藏档案数字资源，汇聚形成档案专题数据库。

b）档案内容提取。围绕特定主题，对馆（室）藏档案内容进行分析，从中采集与主题相关的信息，组织形成档案专题数据库。

c）专项收集归档。围绕专项工作、重大活动、突发事件等，以项目的形式，组织开展相关文件材料收集（征集）、整理、归档、数字化等，形成档案专题数据库。

* + 1. 实施模式

应根据情况，采取独立或合作模式实施建设，可跨区域、跨层级、跨部门合作。具备条件的可引入服务外包，服务外包应符合安全保密工作有关要求。

* + 1. 流程管理

档案专题数据库建设工作主要有选题、数据采集、鉴定与审核、整理与著录、数据清洗、质量检查和验收、查阅利用、管理和维护等业务。应结合实际工作内容，将所涉及业务组织为工作流程，实现各项业务有机衔接，规范管理，确保质量、效率和安全。

* 1. 技术保障
		1. 设备保障

应根据档案专题数据库建设工作需要，配备数据采集处理、存储及备份、安全保密等软硬件，具备条件的可配备档案数据化加工、声像采编等设备。

* + 1. 系统功能

档案专题数据库支撑系统的功能要求参照GB/T 39784，并支持多来源数据整合、多平台发布利用。

* + 1. 技术应用

可应用人工智能、大数据、大模型等技术提高数据采集、开发、利用的效率和精准度，可应用数据可视化技术增强展示的直观性和体验性。

* 1. 选题与建设方案
		1. 选题提出

承担管理职责的部门应结合工作需要和档案资源实际，提出选题意见，并明确其定位、作用和意义。

* + 1. 选题分析

应从必要性、可行性、效益等方面对选题进行分析。

 a）必要性分析。从收集整合档案资源、深化档案资源开发、促进社会各方面利用等方面进行分析，评估选题的必要性。

b）可行性分析。调研与主题相关档案的数量、类型、价值、密级、保管单位、应用场景等情况，了解相关文献资料、研究成果等状况，了解专项工作、重大活动、突发事件的工作安排、进展和档案管理情况，结合政策法规、知识产权保护、安全保密等要求，以及经费、技术、人才保障等因素进行可行性分析。

c）效益分析。对效益指标进行分析，内容包括但不限于经济、社会、文化、生态、可持续影响等指标，预期成果应符合各项效益指标要求。

* + 1. 确定选题

应根据选题分析，对选题进行优化、确定选题。涉及资金预算需求的，应按规定报相关部门审核立项。可采用专家论证方式辅助确定选题。

* + 1. 建设方案

档案专题数据库建设方案应作为档案专题数据库建设工作组织管理、监督指导和实施的依据。

应在深入调研的基础上，本着立足实际、维护历史真实面貌和便于实施的原则，编制档案专题数据库建设方案。

档案专题数据库建设方案主要包括技术方案、业务规范、工作流程、安全保密管理，以及组织实施、经费、人员、制度保障、查阅利用、宣传推广等内容。

* 1. 数据采集
		1. 采集方案

应制定数据采集方案，规范数据采集的内容、范围、数量及质量要求，明确数据采集方式、归集责任与要求等。

* + 1. 采集方式

可采用以下几种采集方式，在线或离线采集数据：

a）对接业务系统，通过接口接收业务系统产生的数据，实现主题相关数据采集。

b）与目标数据库对接，通过数据复制、数据映射或数据导入等，实现主题相关数据采集。

c）经合规授权，通过网络收集有关用户上传数据，或从网站上主动采集主题相关数据。

d）人工或智能技术辅助采集档案中与主题相关的数据。

* + 1. 采集要求

应对数据采集人员进行业务和技术培训，确保按照采集方案确定的采集范围和要求实施数据采集。所采集的数据应来源可靠、内容完整准确、组件齐全、采集过程规范。

* + 1. 数据归集

承担数据采集工作的部门，应按照数据采集方案规定的时间、频次和方式、手续等要求，向指定的部门归集数据。

* 1. 鉴定与审核
		1. 鉴定

应从与主题契合度、来源可靠性、典型性、代表性和重要性等方面对所采集的数据进行鉴定，剔除不符合要求的数据。

* + 1. 审核

应参照档案开放的有关规定，对通过鉴定的数据进行审核。重点审查及标记涉密、敏感及隐私信息，确保在使用中不损害国家安全、社会秩序和公共利益，不侵犯他人合法权益。

* 1. 整理与著录
		1. 整理

数据整理

数据整理应符合以下要求：

a）电子档案应符合真实性、完整性、可用性和安全性要求，格式、元数据、档案信息包等符合国家和档案行业各类别电子文件归档、电子档案管理的标准规范要求。

b）传统载体档案应符合国家和档案行业相应类别文件材料归档管理的标准规范要求，其数字化应符合DA/T 31、DA/T 43、DA/T 62、 DA/T 89等要求。

c）声像档案应主题鲜明，图像画面完整、端正，声音和图像画面清晰，尽可能保证较高的技术质量；除不影响真实性的格式转换、裁剪外，不进行其他技术处理；数码照片如非必要，不能是缩略图。

数据组织

应考虑档案专题数据库的管理和利用需求，按门类、内容、时间、来源等对数据进行分类，实现分散、多元、异构的数据有序组织。

数据编号

应制定数据编号规则，对数据以件为单位进行编号。数据编号的参考格式为“档案专题数据库名称代码-分类代码-顺序号”或“档案专题数据库名称代码-顺序号”，中间用“-”（连接符）隔开（见示例）。档案专题数据库名称代码、分类代码可用英文字母、中文或阿拉伯数字标识；顺序号用阿拉伯数字标识，应按照档案数量最多的类别中文件数确定其位数，位数不足的，前面用“0”补足。

示例1：

南亚博览会档案专题数据库，按年度分类，2024年第105件，编号为：NYBLH-2024-00105。

示例2：

劳模档案专题数据库，不分类，第148件，编号为：劳模-00148。

* + 1. 著录

一般要求

以专项收集归档方式建设档案专题数据库，应按照DA/T 18、DA/T 46、DA/T 54、DA/T 63等档案标准规范实施著录；以档案汇集、档案内容提取等方式建设档案专题数据库，可根据建库需求，从有关档案标准规范设置的著录项目中，选择所需项目实施著录。

应根据需要扩展设置著录项目，对主题、地名、人名、其他专门信息，以及来源信息和使用限制信息进行著录。扩展设置的著录项目应制定著录规则，宜对著录内容的描述语言进行规范，便于数据的聚合和可视化呈现。扩展著录示例见附录A。

主题信息著录

每个档案专题数据库应建立主题词（关键词）表，主题词（关键词）总数宜不超过30个。应根据每件档案内容，选取主题词（关键词）进行著录。多个主题词（关键词）可并列标识，并用“；”（中文分号）分隔。

地名信息著录

根据档案内容，对与主题相关的事件、人物所涉及的地名进行著录。按“国别—行政区划”分级著录。地名用全称或规范简称，在同一个专题数据库内，同一地名的著录信息应一致。多个地名应逐个著录，用“；”（中文分号）分隔。历史地名照原文著录，可将考证后的当今地名附后注明。

人名信息著录

选择与主题内容直接相关的人物，按其在档案中呈现的姓名与身份，规范进行人名信息著录。可根据需要设置多个著录项目，著录姓名、籍贯、职衔或身份、证件号码等信息。

其他专门信息著录

应对反映主题、数据管理和提供利用有重要作用的其他专门信息进行著录。

来源信息著录

应对所采集数据的来源信息进行著录，示例见附录B。来源信息著录应符合以下要求：

a）来源信息应反映提供数据的机构或个人信息及获取方式；

b）以档案汇集、档案内容提取方式获取的数据，应著录原档号信息；

c）以专项收集归档方式获取的数据，应著录移送者信息。

使用限制信息著录

 应根据有关政策规定、法律法规、协议或移交记录，著录数据使用限制信息，示例见附录C。

通用著录项目对应关系

可根据跨库、跨类检索利用的需要，明确有关著录项对应关系，示例见附录D。

* 1. 质量管理与验收
		1. 数据清洗

根据需要，选择在数据采集、鉴定、整理、著录、质量评估等环节进行数据清洗。

应对数据缺失、重复数据、多余字符、乱码、逻辑错误等著录不规范问题，以及电子文件损坏、格式不规范、与目录不对应等异常情况进行检核修正。

开展数据清洗，应制定操作规程，明确数据清洗规则、流程、方法和要求，记录数据清洗过程；每次进行数据清洗工作前，应进行数据备份。

* + 1. 数据质量检查

在数据整理、著录、清洗等数据处理工作环节，应按照数据采集方案的要求进行数据质量检查，合格后方可进入下一工作流程。

* + 1. 验收

档案专题数据库建设工作完成后，经测试和试运行，承担管理职责的部门可依据建设方案、数据采集方案等，组织开展专题数据库建设工作验收。

* 1. 查阅利用与宣传推广
		1. 查阅

档案专题数据库查阅利用参照档案查阅利用制度执行。档案专题数据库共享按照档案专题数据库建设方案的安排或共享协议实施。

* + 1. 公开发布

档案专题数据库公开发布应遵守有关档案公布、信息上网等制度；应根据安全保密要求、数据使用限制要求等，合理选择数据向有关发布利用平台或渠道分发投放，必要时应进行脱敏脱密处理。

* + 1. 查阅利用平台

应建立档案专题数据库查阅利用平台，包括但不限于档案查阅利用系统、多媒体交互展示系统、政务信息共享系统、网站系统、社交媒体等。

* + 1. 产品开发

可开发档案专题数据库多媒体光盘、电视专题片、短视频、电子像册等产品，多形式推广利用。

* + 1. 宣传推广

档案专题数据库宣传推广工作宜与建设工作同步筹划；宜利用电视、广播、报刊、网络、融媒体等，多渠道进行宣传。宜配合档案专题数据库发布，举办形式多样的宣传推广活动，制作宣传手册、音像制品和文创产品等宣传品。

* 1. 管理与维护
		1. 反馈评价

档案专题数据库查阅利用期间宜对用户反馈的意见进行收集，对档案专题数据库的用户满意度、成本效益、可持续性、安全性等绩效指标进行综合评价。

* + 1. 动态管理

应根据用户反馈意见和综合评价结果，不断进行完善；根据专题内容、性质及馆（室）藏档案增长变化情况，适时对数据库进行扩展、更新。

* + 1. 维护备份

档案专题数据库建设完成后，应纳入馆（室）档案资源安全体系，开展日常管理维护，参照DA/T 99做好备份工作。

1.
2. （资料性）
扩展著录（示例）

以下给出了扩展著录的示例。

|  |  |  |
| --- | --- | --- |
| 扩展著录项目名称 | 内容 | 著录示例 |
| 主题信息 | 根据内容和主题归属，选取主题词（关键词）进行著录 | 1.反映数据采集主题的主题词（关键词）：脱贫攻坚，乡村振兴…2.反映具体活动实际的主题词（关键词）：基础设施，产业发展，保障措施… 3.反映事物组成部分的主题词（关键词）：传统民居，宗教建筑，古村落…  |
| 地名信息 | 对与主题相关事件、人物所涉及的地名进行著录 | 1.XX省XX市XX区XX街道XX社区2.XX省XX市3.云南省六顺县（云南省普洱市思茅区） |
| 人名信息 | 选择与主题内容直接相关的人物，按其在档案中呈现的姓名和身份，规范进行人名信息著录 | 1.姓名：张三2.身份：XX县县长，XX局局长，养殖大户，非物质文化遗产传承人…3.籍贯：XX省XX市… |
| 其他专门信息 | 对反映主题、数据管理和提供利用有重要作用的其他专门信息进行著录 | 1.民族：傣族，景颇族…2.区域：革命老区，边境地区…3.类别：建筑，交通，设施，政务… |
|  注：地名信息著录示例3中的“云南省六顺县”为历史地名，“云南省普洱市思茅区”为考证后的当今地名 |

1. （资料性） 来源信息著录（示例）

以下给出了来源信息著录的示例。

|  |  |  |
| --- | --- | --- |
| 来源信息著录项 | 内容 | 示例 |
| 来源名称 | 反映提供数据的机构或个人信息 | 1.XX省档案馆2.XX省发展与改革委员会3.XX卫视4.张三，XX村农民5.“http://www.yunnan.cn”，XX网 |
| 获取方式 | 获取数据的途径 | 接收，XX馆（室）藏，获赠，购买，寄存，下载，收录，其他… |
| 移送者 | 反映专项收集归档方式中数据移送者信息 | 1.XX组委会2.XX指挥中心3.XX省应急管理厅 |

1. （资料性） 使用限制信息著录（示例）

以下给出了使用限制信息著录的示例。

|  |  |  |
| --- | --- | --- |
| 使用限制信息著录项 | 内容 | 示例 |
| 控制标识 | 根据数据安全利用需要设定的相应管理标识 | 公开，档案馆内部使用，档案涉及人员或机构使用，政务网使用，互联网公开… |
| 授权行为 | 需经有关单位批准、或需著作权、肖像权使用授权等限制使用要求 | 报批有关单位使用，作者授权使用，肖像授权使用… |
| 密级 | 反映数据保密程度 | 非密、工作秘密、秘密，机密，绝密 |

1. （资料性） 通用项目对应关系表（示例）

以下给出了通用项目对应关系示例。

|  |  |  |  |
| --- | --- | --- | --- |
| 通用项目 | 文书类档案元数据 | 照片类档案元数据 | 录音录像类档案元数据 |
| 编号 | 元数据 | 编号 | 元数据 | 编号 | 元数据 |
| 题名 | M22 | 题名 | M6 | 题名 | M7 | 题名 |
| 档号 | M8 | 档号 | M5 | 档号 | M6 | 档号 |
| 档案馆代码 | M4 | 档案馆代码 | M1 | 档案馆代码 | M1 | 档案馆代码 |
| 内容描述 | M21 | 内容描述 | M15 | 主题 | M19 | 主题 |
| 来源 | M2 | 来源 | M29 | 来源 | M23 | 来源 |
| 格式信息 | M47 | 格式信息 | M51 | 格式信息 | M48 | 格式信息 |
| 责任者 | M32 | 责任者 | M7 | 责任者 | M8 | 责任者 |
| 日期 | M33 | 日期 | M11 | 时间 | M13 | 时间 |
| 密级 | M38 | 密级 | M35 | 密级 | M29 | 密级 |
| 保管期限 | M12 | 保管期限 | M33 | 保管期限 | M27 | 保管期限 |

参考文献

1. GB/T 5271.1-2000 信息技术 词汇 第1部分:基本术语
2. GB/T 20273-2019信息安全技术 数据库管理系统安全技术要求
3. GB/T 18894 电子文件归档与电子档案管理规范
4. GB/T 35274 数据安全技术大数据服务安全能力要求
5. GB/T 39362 党政机关电子公文归档规范
6. GB/T 42727 政务服务事项电子文件归档规范
7. DA/T 2 科学技术研究项目档案管理规范
8. DA/T 13 档号编制规则
9. DA/T 22 归档文件整理规则
10. DA/T 32 公务电子邮件归档管理规则
11. DA/T 50 数码照片归档与管理规范
12. DA/T 70 文书类电子档案检测一般要求
13. DA/T 78 录音录像档案管理规范
14. DA/T 80 政府网站网页归档指南
15. DA/T 88 产品数据管理(PDM)电子文件归档与电子档案管理规范
16. DA/T 103 招标投标电子文件归档规范
17. DA/T 104 ERP系统电子文件归档和电子档案管理规范
18. 网络数据安全管理条例（中华人民共和国国务院令第790号）
19. 重大活动和突发事件档案管理办法（国家档案局令第16号）
20. 国家档案馆档案开放办法（国家档案局令第19号）
21. 电子档案管理办法（国家档案局令第22号）
22. 国家档案局办公室关于印发《电子档案管理系统基本功能规定》的通知（档办发〔2017〕3 号）

