

考试资料职业发展
技术读书笔记分享

B站/闲鱼：大西洋活跃的锅巴
公众号：不太甜

DAMA-DMBOK

数据管理知识体系指南CDGA/CDGP认证

第9章文件和内容管理（完整课程视频请扫描二维码）



第9章 文件和内容管理



01

引言

02

活动

03

工具和方法

04

实施指南

05

数据集成和互操作治理



01

引言

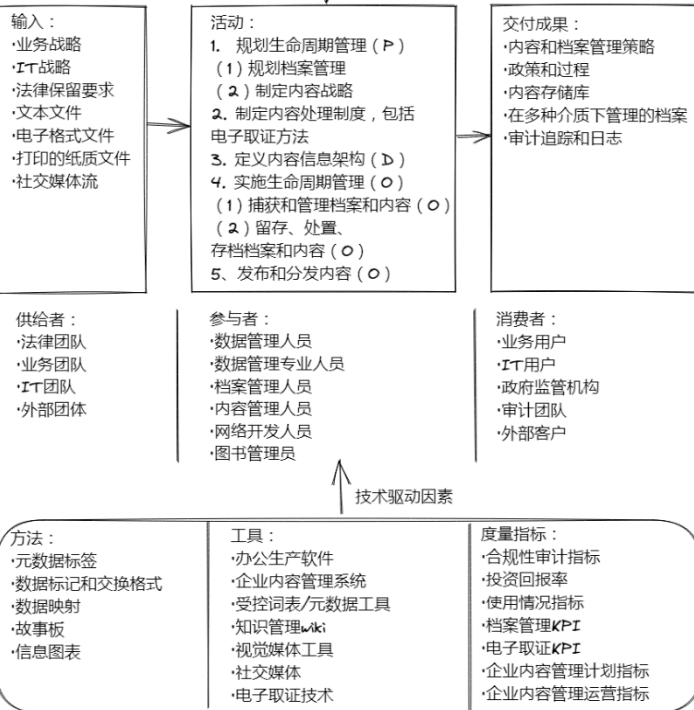
定义、业务驱动因素、目标和原则、基本概念

定义：对任何形式或媒介的数据及信息进行生命周期管理的计划、实施和控制活动。

目标：

- 履行与档案管理有关的法律义务并达到客户的期望
- 确保能够高速有效的存储、检索、使用文件和内容
- 确保结构化和非结构化内容之间的集成能力

业务驱动因素



技术驱动因素

(P) 计划 (C) 控制 (D) 开发 (O) 运营

语境关系图：文件和内容管理

文件和内容管理是针对存储在关系型数据库之外的数据和信息的采集、存储、访问和使用过程的管理。重点在于保持文件和其他非结构化或半结构化信息的完整性，并使这些信息能够被访问。

主要业务驱动因素包括法规遵从性要求、诉讼响应能力和电子取证请求能力以及业务连续性要求。考题提高效率是改进文件管理的驱动力。

目标：

- 1) 确保能够高效地采集和使用非结构化的数据和信息
- 2) 确保结构化和非结构化数据之间的整合能力
- 3) 遵守法律义务并达到客户预期

原则：

- 1) 组织中的每个人都应该在保护组织的未来方面发挥作用。
- 2) 档案和内容处理方面的专家应充分参与制度和规划的制定

档案保存指导原则（GARP）

- 1) 问责原则
- 2) 完整原则
- 3) 保护原则
- 4) 遵从原则
- 5) 可用原则
- 6) 保留原则
- 7) 处置原则
- 8) 透明原则

1、内容

是指文件、档案或网站内的数据和信息，通常基于文件所代表的概念以及文件的类型或状态来管理。

(1) 内容管理

包括用于组织、分类和构造信息资源的流程、方法和技术，以便以多种方式存储、发布和重复使用这些资源。

(2) 内容元数据

元数据对于管理非结构化的数据至关重要，无论是传统上认为的内容和文件，还是现在理解的“大数据”。如果没有元数据，就无法对内容进行编目和组织。非结构化数据内容的元数据基于：

- | | |
|----------|---------|
| 1) 格式 | 2) 可搜索性 |
| 3) 自我描述性 | 4) 既有模式 |
| 5) 内容主题 | 6) 需求 |

1、内容

(3) 内容建模

内容建模是将逻辑内容概念转换为具有关系的内容类型、属性和数据类型的过程。属性描述关于该内容的一些特定的和可区分的信息；数据类型限制了属性可以包含的数据的类型，从而使验证和处理成为可能。元数据管理和数据建模技术用于开发内容模型。内容建模有两个层次：

成果
第一是信息产品级别，产出一个像网站一样的实际可交付

第二是组件级别，进一步详细说明了构成信息产品模型的元素，模型中的详细程度取决于重用和构造所需的粒度。

1、内容

(4) 内容分发方法

1) 推式

在推式系统中，用户按照预先确定的时间表选择传送给他们的内容类型。简易信息聚合RSS。

2) 拉式

用户通过互联网获取内容

3) 交互式

第三方电子销售点EPOS的应用或面向客户的网站，需要在企业应用之间交换大量的实时数据。应用程序之间共享数据的选项，包括企业应用程序集成（EAI）、更改数据采集、数据集成和EII

2、受控词表

是被明确允许用于通过浏览和搜索对内容进行索引、分类、标引、排序和检索术语的定义列表。系统地组织文件、档案和内容离不开受控词表。

(1) 词汇表管理

提高信息存储和检索系统、web导航系统和其他环境的有效性手段，这些环境通过使用语言的某种描述来识别和定位所需的内容。控制词汇表的主要目的是实现采用一致并便于检索的方式描述内容对象

针对任何给定的词汇进行定义、寻源、导入和维护的过程。

(2) 词汇表视图和微控制词汇表

词汇表视图是受控词表的子集，涵盖了受控词表领域内有限范围的主题。

微控制词汇表是包含一般词汇表中不包含的高度专业化术语的词汇表。

(3) 术语和选择列表

术语列表仅仅就是一个列表

选择列表通常隐藏在应用程序中

(4) 术语管理

术语管理包括在受控词汇表中建立术语之间的关系，关系类型可以分为以下3种：

1) 等价术语关系

术语映射功能

2) 层次化关系

广义到狭义或整体到部分

3) 关联关系

相关联但非层次化

(5) 同义词环和规范表

同义词环是指一组含义大致相同的术语。允许搜索其中一个术语的用户去访问与该词环其他术语相同的内容。目的是用于检索而非索引。

规范表是描述性术语的受控词汇表，旨在促进特定领域或范围内的信息检索。其术语处理方法和同义词环不同

(6) 分类法

是指任何分类或受控词表的总称。最著名的是林奈开发的生物分类系统。分类法有不同的结构：

- 1) 扁平分类法
- 2) 层次分类法：地理信息，从大陆到街道
- 3) 多重层级结构：它是具有多个节点关系规则的树状结构
- 4) 面分类法：它指的是每个节点与中心节点相关联，其形状看起来像星型图。
- 5) 网状分类法。

(7) 分类方案和打标签

分类方案是代表受控词表的代码

大众分类法是通过社交标签对在线内容术语和名称分类的方案

(8) 主题词表

又称叙词表，是一种用于内容检索的受控词表。

(9) 本体

代表一套概念和它们在某个领域内概念之间的关联。

分类法和本体的两个主要区别：

- 1) 分类法为给定的概念区域提供数据内容分类
- 2) 在分类法或数据模型中，定义是已知的，别无其他，称为封闭世界假设。在本体中，基于现存的关系可以推断出可能的关系。因此未明确声明的内容也可能是真的，称之为开放世界假设。

避免落入导致歧义和混淆的常见陷阱，这类情况主要有：

- 1) 无法区分实例关系和子类关系
- 2) 将事件建模为关系
- 3) 术语缺乏清晰度和独特性
- 4) 将角色建模为类
- 5) 无法重复使用
- 6) 混淆建模语言的语义和概念
- 7) 使用基于网络的、跨平台的工具（如OOPS）进行本体验证有助于陷阱的诊断和修复。

文件是包含任务说明，对执行任务或功能的方式和时间的要求以及任务执行和决策的日志等的电子或纸质对象。

只有部分文件才能称为档案。档案可用于证明所做的决策和所采取的行动是符合程序的；可作为组织业务活动和法规遵从的证据。

(1) 文件管理

文件管理包括在文件和档案的整个生命周期中控制和组织它们的流程、方法和技术。它包括电子和纸质文件的存储、编目和控制。

- 1) 编目
- 2) 制度
- 3) 分类
- 4) 存储
- 5) 检索和流转
- 6) 保存和处置

(2) 档案管理

是文件管理的一部分，管理档案有一些特殊的要求，包括整个档案的生命周期：从档案的创建或接收到处理、分发、组织和检索，再到处置。

精心管理的档案具有以下特点：

- 1) 内容
- 2) 背景
- 3) 及时性
- 4) 永久性
- 5) 结构

(3) 数字资产管理

4、数据地图

是所有ESI数据源、应用程序和IT环境的清单，其中包括应用程序所有者、保管人、相关地理位置和数据类型等信息。

5、电子取证

早期案例评估

早期数据评估

6、信息架构

- 1) 受控词表
- 2) 分类法和本体
- 3) 元数据映射
- 4) 搜索功能规格
- 5) 用例
- 6) 用户流

7、搜索引擎

是一种根据属于搜索信息并检索内容中包含这些术语网站的软件。

包含几个组件：

适当的搜索引擎软件、漫游网络的爬虫、将找到内容的统一资源定位符（URL）保留起来的存储、索引遇到的关键字和文本以及排名规则等。

8、语义模型

是一种知识建模，描述一系列概念网络以及它们之间的关系。

包含语义对象和语义约束。

9、语义搜索

侧重于语义和语境而非预先设定的关键字。

10、非结构化数据

11、工作流

内容管理系统（CMS）实现自动化



02

活动

规划、创建、定义、实施、发布和分发

1、规划生命周期的管理

文件生命周期的规划：从文件的创建或接收文件后的分发、存储、检索归档和潜在的销毁。规划包括开发分类/索引系统和分类法，以实现文件的存储和检索。

- 1、规划档案管理
- 2、制定内容策略

文件管理制度

- 1) 审计的范围和合理性
- 2) 重要档案的鉴定和保护
- 3) 保留档案的目的和保管期限表
- 4) 如何响应信息保留命令，即针对已过保留期的诉讼信息要求
- 5) 本地和异地存储档案的要求
- 6) 硬盘驱动器和共享网络驱动器的使用和维护
- 7) 对电子邮件管理，从内容管理的角度进行处理
- 8) 合理的档案销毁方法，如预先批准的供应商和销毁证明的收据

- (1) 社交媒体制度
- (2) 设备访问制度
- (3) 处理敏感数据
- (4) 应对诉讼

3、定义内容信息架构

使用基于内容的索引或元数据搜索

基于用户的需求和偏好，索引的设计侧重于索引的关键或属性的决策选项。

确保构建受控词汇表、索引、信息检索分类方案工作的互相协调。

4、实施的生命周期管理

1、获取档案和内容

2、管理版本的控制

1) 正式控制

2) 修订控制

3) 托管控制

3、备份和恢复

4、管理保管和处置

没有优先删除无附加值的信息是因为：

1) 制度不适用

2) 对某一个人来说是无附加值的信息，但对另一个人来说却是有价值信息

3) 无法预见当前的无附加值实体和/或电子档案未来可能的需求

4) 对档案管理的不认可

5) 无法决定删除哪些档案

6) 做决定与移除实体和电子档案的感知成本

7) 电子空间很便宜，购买更多的空间比归档和移除过程更容易

5、审计文件和档案

1) 定义组织驱动因素

2) 一旦确定检查测量的内容和要使用的工具，就开始收集相关数据

3) 报告结果

4) 制定下一步的行动计划和时间表

5、发布和分发内容

发布和分发内容

- 1、开放访问、搜索和检索
- 2、通过可接受的渠道分发



03

工具和方法

1、文件管理

基于不同规则的工作流：1) 手动工作流 2) 基于规则的工作流 3) 动态规则
允许基于内容的不同工作流。

(1) 数字资产管理： 数字资产的编目、存储和检索

(2) 图像处理系统： 用于采集、转换和管理纸质件的影像和电子文件，采集使用的技术包括扫描、光识别和智能字符识别或表单处理。

光学字符识别OCR

智能字符识别ICR是一种更先进的可以处理打印和草书手写的文本

(3) 档案管理系统： 可能要有自动保管和处理、电子取证支持和长期归档的能力。支持重要的档案程序，以保留关键业务档案，可以与文件管理系统集成。

2、内容管理系统

content management system,CMS .用于收集、组织、索引和检索内容，将内容存储为组件或整个文件，同时保持组件之间的链接。还可以提供修改文档内容的控件。

3、内容和文件工作流

工作流工具支持业务流程、路由内容和文件、指派工作任务、跟踪状态以及创建审计跟踪。在内容发布之前，工作流应支持对内容的审核和批准。

协作工具

可用于收集、存储、工作流程和管理与团队活动相关的文件。社交网络可以使个人和团队在内部共享文件和内容，并通过博客、**wikis**、**RSS**和标签与外部团队进行交互。

受控词汇表和元数据工具

- 1) 在组织中被用作数据指南的数据模型
- 2) 文件管理系统和办公软件
- 3) 元数据库、术语表或目录
- 4) 分类法和分类法之间的交叉参考模式
- 5) 集合索引、文件系统、民意调查、档案、位置或异地控股
- 6) 搜索引擎
- 7) 非结构化数据的BI工具
- 8) 企业和部门同义词表
- 9) 已发布的报告库、内容和参考书目及其目录。

1、可扩展标示语言（XML）

1) XML提供了将结构化数据整合到具有非结构化数据的关系型数据库中的功能

2) XML可以将结构化数据与文件、报表、电子邮件、图像、图形、音频和视频文件中的非结构化数据集成在一起。

3) XML还可以用于建立企业或公司门户网站（B2B、B2C），为用户提供了一个可以连接到各种内容的站点。

4) XML可以识别和标记非结构化数据/内容，以便计算机应用程序可以理解并处理它们

2、基于JavaScript的轻量级数据交换格式JSON

是一种开放的、轻量级的数据交换标准格式。它的文本格式独立于语言，易于解析，但仍是使用C语言家族的习惯方式。JSON有两个结构：一是无序“名称/值对”集合，也被称为对象；二是值的有序集合，也被称为数组。

3、资源描述框架RDF和相关的万维网联盟W3C规范

RDF——主谓宾

网络本体语言（W3C Web Ontology Language，OWL）是RDF的词汇表扩展

简单知识组织系统（Simple Knowledge Organization System，SKOS）是一个

RDF词汇表

4、Schema.org:

提供了一组用于页面标记的共享词汇表或模式，以便主流的搜索引擎可以理解它们。侧重于网页上的文字含义以及术语和关键词。

词汇表集合还可用于结构化数据的互操作（如与JSON）

电子取证技术提供了许多功能和技术，如早期案例评估、收集、辨认、保全、处理、光学字符识别（OCR）、剔除、相似性分析和电子邮件线程分析。技术辅助审查（TAR）是一个工作流或过程。在此过程中，团队可以审查选定的文件并标记相关与否。

1、诉讼应诉手册

指引手册汇编内容包括：

- 1) 给每个部门建立制度和过程清单
- 2) 为各主题起草制度，如诉讼保留、文件保管、归档和备份
- 3) 评估IT工具的能力，如电子取证索引、搜索和搜集、数据隔离和保护工具以及非结构化ESI源/系统
- 4) 识别和分析相关的法律问题
- 5) 制定沟通和培训计划，根据预期对员工进行培训
- 6) 确定可以提前准备的材料，以便根据法律案例进行裁减
- 7) 在需要外部服务的情况下分析供应商服务
- 8) 制定如何处理通知的流程并保持指引手册的最新状态

2、诉讼应诉数据映射

数据映射是一个信息系统的目录，描述了系统以及系统的用途、包含的信息、保留策略和其他特征。从目录中通常会识别档案系统、源应用系统、存档、灾难恢复副本或备份以及用于每个系统的介质。



04

实施指南

就绪评估/风险评估、组织和文化变革

1、就绪评估/风险评估

ECM就绪评估的目的是确定内容管理需要改进的方面，以及组织对改变其流程以满足这些需求的适应程度。数据管理成熟度评估模型可以在此过程中提供帮助。

具体的ECM关键成功因素包括对现有内容的审核和分类、合适的信息体系结构、内容生命周期的支持、适当元数据标签的定义以及在ECM解决方案中自定义的功能的能力。

（1）档案管理成熟度

ARMA的公认档案保管原则：

- 1）1级，低于标准的
- 2）2级，发展中的，认识到信息治理和档案保管可能对组织产生的影响
- 3）3级，基本的，必须满足法律法规的最低要求
- 4）4级，积极的，已经建立了一个以持续改进为重点的主动式信息治理计划
- 5）5级，完成变革的。信息治理已经融入企业的基础架构和业务流程中。

（2）电子取证评估

2、组织和文化变革

理想的解决方案是集中和安全管理单一存储库，在整个企业中实施明确定义的策略和流程。



05

文件和内容治理

治理架构和度量指标

1、信息治理架构

驱动因素包括：

- 1) 法律和法规遵从
- 2) 档案的合理处置
- 3) 对电子取证的前瞻性准备
- 4) 敏感信息的安全
- 5) 电子邮件和大数据等风险领域的管理

1、信息治理架构

GARP原则：

- 1) 为问责制度提供高层支持
- 2) 教育员工了解信息治理制度
- 3) 根据正确的档案代码或分类类目对信息进行分类
- 4) 确保信息的真实性和完整性
- 5) 确保官方档案是电子档案，除非另有规定
- 6) 制定制度，使业务系统和第三方信息与信息治理标准保持一致
- 7) 存储、管理、访问、监控和审核已批准的企业存储库以及档案和内容系统
- 8) 保护机密或个人身份信息
- 9) 控制不必要的信息增长
- 10) 在信息的生命周期结束时对其进行处置
- 11) 遵守信息的请求（如取证、传票等）
- 12) 持续改进

信息治理参考模型（IRGM）

展示了信息治理与其他组织功能的关系。外环包括了实施信息管理过程中涉及的制度、标准、流程、工具和基础架构等方面，中心展示了一张生命周期图。

2、信息的激增

一般来说，非结构化数据的增长速度远远快于结构化数据，这增加了治理的难度。

3、管理高质量的内容

- 1) 生产者
- 2) 消费者
- 3) 时间
- 4) 格式
- 5) 分发

1、档案管理

衡量档案管理实施成功的标准：

- 1) 每个用户有百分之几的文件和电子邮件被认定为是公司的档案
- 2) 被认定为公司档案中有百分之几得到了控制
- 3) 在所有存储的档案中有百分之几的档案应用了适当的保管规则

2、电子取证

电子取证常见的一个KPI指标是成本降低，另一个KPI指标是对比被动收集信息的方式，提前收集信息时提高效率。还有一个是组织可以快速地进行法律保留通知的过程

主要集中在的7个方面：活动、监护人、系统、介质、状态、格式和质量保证。

3、企业内容管理

指定一些与业务解决方案的价值相关的KPI，如：

1) 财务KPI可以包括ECM系统的成本、与物理存储相关的成本降低以及运营成本下降的百分比

2) 客户KPI可以包括首次与客户联系时就解决了他的问题的比例和客户投诉的数量

3) 代表更有效和更高效的内部业务流程的KPI可以包括在使用工作流程和过程自动化后。文书工作减少的百分比以及错误减少的百分比

4) 培训KPI可包括管理和非管理培训课程的数量

5) 风险缓解KPI可以包括取证成本的降低以及追踪电子取证请求的审计跟踪数量。

1、规划指标：ECM项目数、采用率和用户满意度

2、运营指标：停机时间、用户数量

3、特定ECM指标：存储利用率（如使用企业内容管理前后的存储容量对比）、检索性能

4、信息检索能力：

查准率：检索到的文件与实际相关的比率

查全率：检索到的涉及所有相关文档的比率

考试资料职业发展
技术读书笔记分享

B站/闲鱼：大西洋活跃的锅巴
公众号：不太甜

本章完结 感谢观看

完整课程视频请扫描二维码咨询

