



普通高校“十三五”规划教材·信息管理与信息系统系列

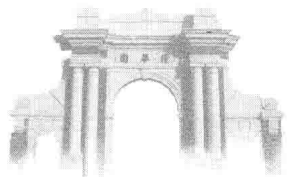
# 电子文件管理



王 英  
蔡盈芳 ◎ 主编  
黄 磊

清华大学出版社

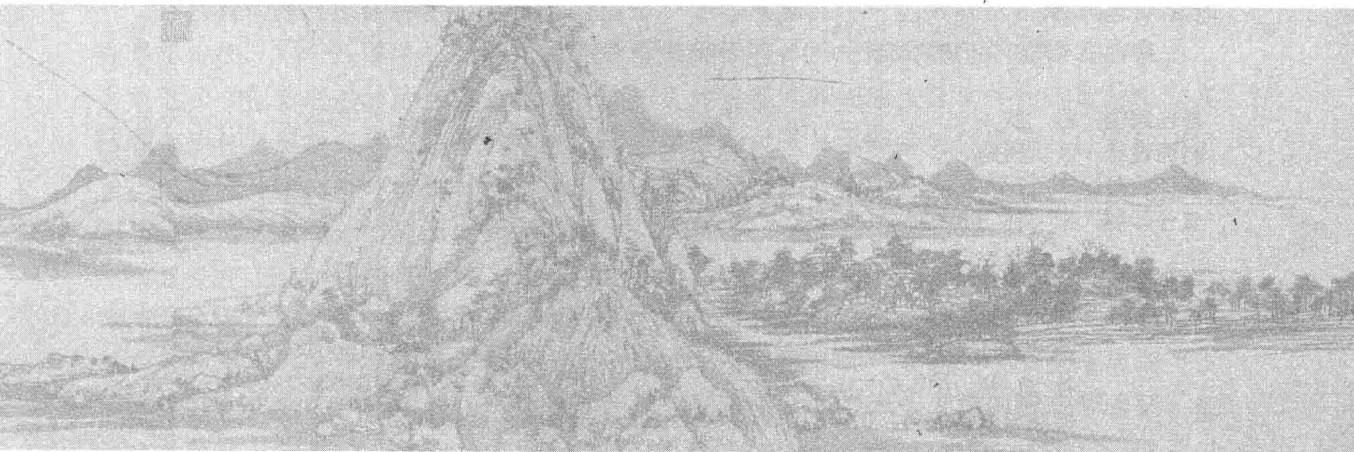




普通高校“十三五”规划教材·信息管理与信息系统系列

# 电子文件管理

王英 蔡盈芳 黄磊 ◎ 主编



清华大学出版社  
北京

## 内 容 简 介

在知识经济背景下,电子文件呈现出和传统实物文件显著不同的诸多新特征,知识成为电子文件的重要属性。电子文件管理从传统的档案实体管理向知识管理演变,信息资源管理、信息检索、知识管理等信息科学领域的内容逐步成为电子文件管理中必须关注的重要主题。

本书基于对知识经济时代电子文件管理新特征的分析,全面介绍了电子文件管理的基本理论和方法。全书分为8章,包括电子文件和电子文件管理的基本概念、电子文件管理基础理论、电子文件全生命周期管理流程、元数据、电子文件知识库管理、电子文件知识分析、电子文件网络信息资源管理和电子文件管理趋势与政策。

本书可作为图书情报、档案管理、信息管理与信息系统、管理科学与工程、保密管理等专业的本科生教材,也可作为档案管理相关技术人员和管理干部的参考资料。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13801310933

### 图书在版编目(CIP)数据

电子文件管理/王英,蔡盈芳,黄磊主编. —北京:清华大学出版社,2016  
(普通高校“十三五”规划教材.信息管理与信息系统系列)  
ISBN 978-7-302-45318-5

I. ①电… II. ①王… ②蔡… ③黄… III. ①电子档案—档案管理—高等学校—教材  
IV. ①G275.7

中国版本图书馆CIP数据核字(2016)第256064号

责任编辑:彭欣

封面设计:汉风唐韵

责任校对:宋玉莲

责任印制:宋林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, [c-service@tup.tsinghua.edu.cn](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课件下载: <http://www.tup.com.cn>, 010-62770175 转 4506

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:185mm×260mm

印 张:17.5

字 数:403千字

版 次:2016年12月第1版

印 次:2016年12月第1次印刷

印 数:1~4000

定 价:37.00元

产品编号:065918-01

# 前言

在当前信息化和知识经济不断升温的浪潮中，电子文件大量生成和普遍应用，已经成为经济社会的基本信息支撑。1996年，在北京召开的第十三届国际档案大会为我国的档案学者打开了电子文件管理这一新的研究视野，电子文件管理得到了我国档案界学者和档案行政管理部门的高度重视。我国先后制定实施了一系列电子文件管理的规范指导文件，电子文件管理在国家战略层面得到了全面规划和积极推进。

随着数字信息技术的快速发展，电子文件的内涵也愈加丰富，往往是由多种不同类型的计算机文件通过链接、嵌入或其他方式组合的复合体，形成电子文档，并通过归档保存形成电子档案。因此，电子文件呈现和传统实物文件显著不同的诸多新特征，知识成为电子文件的重要属性。在信息化背景下，传统基于手工档案管理实践经验的档案学理论面临诸多新的挑战。电子文件管理带来了档案学基础理论的冲击，电子文件管理从传统的档案实体管理向知识管理演变，信息资源管理、信息检索、知识管理等信息科学领域的研究内容逐步成为电子文件管理中必须关注的重要主题。

在新的信息技术条件下，档案学与信息科学逐渐呈现学科融合现象，电子文件呈现许多新特征，对于高等院校的电子文件管理课程内容体系的建设提出了新要求。但是，当前电子文件管理教材的内容主要关注电子文件管理实践，重点是文件的管理策略和文件管理的技术解决方案。尽管近年来我国档案学者对于知识经济时代电子文件的知识价值已经有所关注，但现有教材在知识管理、知识分析和网络信息资源管理等相关内容方面仍嫌薄弱，或仍停留在概念介绍层面。

正是在上述认识的基础上，本书将电子文件管理置于新的知识经济时代的背景下，确定了本书的编写特色，即基于当前档案学与信息科学学科融合的新趋势，将知识管理原理和技术融入电子文件管理理论。基于此，本书以电子文件管理的整个生命周期为主线，介绍知识管理环境中电子文件管理的理论和方法，包括电子文件的概念和特征、电子文件管理基础理论、电子文件全生命周期管理方法、元数据、电子文件知识库管理、电子文件知识分析、电子文件网络信息资源管理和电子文件管理趋势与政策。

全书共分8章。

第1章以数字信息时代对于电子文件管理的影响为出发点，分析了数字信息资源及其特性，数字信息时代的电子文件特征；阐述了电子文件的概念、定义和种类；分析了电子文件的原始性与真实性，分析了电子文件的法律证据效力。

第2章以电子文件管理的基础理论为主线，阐述了文件生命周期理论的产生、发展和基本内容及其意义；阐述了文件连续体理论概念、管理框架及其影响；分析了文档一

体化管理模式及电子文件管理由“保管模式”向“后保管模式”转变的趋势。

第3章以电子文件的全生命周期管理为主线,分析了电子文件的管理原则、管理要求、管理环节和管理流程标准;分析了电子文件的形成管理的要求;并对电子文件归档、存储、保管、利用和处置的整个过程进行了分析。

第4章以电子文件的元数据管理为主线,阐述了元数据的概念由来、内涵和分类,分析了元数据在电子文件中的作用,介绍了国内外主要的元数据标准,分析了电子文件元数据管理的过程,并通过实例对元数据的详细定义方法进行了分析。

第5章以电子文件知识库管理为主线,分析了电子文件管理与知识管理的融合,知识与知识管理的主要概念,着重学习语义知识库技术和基于本体的语义知识库建模方法,并阐释了知识地图的基本原理和构建技术。

第6章以电子文件知识分析为主线,介绍了电子文件文献计量经典定律、基本方法和计量指标,以及常用的文献数据库;在电子文件知识分析的技术方法中,重点阐述了文本挖掘的概念和算法,并介绍了中文分词的算法和系统。

第7章主要涉及电子文件网络信息资源的概念、Web链接结构与信息内容;搜索引擎和网络爬虫的工作原理与相关工具;网络链接分析基本概念和算法;网络引文分析概念和指标;社会网络分析基本概念、描述模型、分析方法和工具。

第8章讨论新技术环境下的电子文件管理趋势和政策。主要分析了新技术条件下电子文件管理面临的风险及其新趋势,讨论了电子合同和电子发票等新型电子文件的管理思路,分析了各主要发达国家和我国的电子文件管理体制及主要的电子文件管理政策法规。

本书由北京交通大学王英、国家档案局蔡盈芳和北京交通大学黄磊主编。详细编写的章节为蔡盈芳、黄磊(第1~3章),蔡盈芳、王英(第4章和第8章),王英(第5~7章)。冉青云、乔壁君同学对本书的案例分折亦有贡献。本书在写作过程中,得到北京交通大学国家保密学院优质课程建设项目资助,在此表示感谢。此外,本书写作中参考和引用了大量文献,在此一并致谢。

由于水平有限,不妥之处在所难免,敬请读者批评指正。

编者

# 目 录

|                          |    |
|--------------------------|----|
| 第 1 章 电子文件概述             | 1  |
| 1.1 数字信息与电子文件管理          | 1  |
| 1.1.1 数字信息资源及其特性         | 1  |
| 1.1.2 数字信息时代的电子文件        | 5  |
| 1.2 电子文件的概念              | 7  |
| 1.2.1 文件、文档和档案           | 7  |
| 1.2.2 电子文件定义和种类          | 8  |
| 1.2.3 电子文件与电子档案          | 12 |
| 1.3 电子文件与电子证据            | 12 |
| 1.3.1 电子文件的原始性与真实性       | 12 |
| 1.3.2 电子文件的法律证据效力        | 14 |
| 本章小结                     | 15 |
| 第 2 章 电子文件管理基础理论         | 16 |
| 2.1 文件生命周期理论             | 16 |
| 2.1.1 文件生命周期理论产生的专业背景    | 16 |
| 2.1.2 中外文件生命周期理论的发展历程    | 17 |
| 2.1.3 文件生命周期理论的基本内容      | 18 |
| 2.1.4 文件生命周期理论的理论意义和实践价值 | 19 |
| 2.2 文件连续体理论              | 20 |
| 2.2.1 文件连续体模式及其概念的形成     | 20 |
| 2.2.2 文件连续体管理框架          | 21 |
| 2.2.3 文件连续体对电子文件理论的影响和价值 | 23 |
| 2.3 文档一体化管理              | 24 |
| 2.3.1 电子文件的“后保管模式”       | 24 |
| 2.3.2 文档一体化管理模式          | 27 |
| 2.3.3 电子文件中心             | 27 |
| 本章小结                     | 29 |
| 第 3 章 电子文件全生命周期管理        | 31 |
| 3.1 电子文件管理流程             | 31 |
| 3.1.1 电子文件管理原则           | 32 |
| 3.1.2 电子文件管理要求           | 34 |

|            |                      |           |
|------------|----------------------|-----------|
| 3.1.3      | 电子文件管理环节             | 38        |
| 3.1.4      | 电子文件管理流程标准           | 39        |
| 3.2        | 电子文件形成管理             | 43        |
| 3.2.1      | 管理过程能保证电子文件的真实、可靠、安全 | 43        |
| 3.2.2      | 电子文件形成时按一定格式存储       | 44        |
| 3.2.3      | 生成并能够捕获相应的元数据        | 44        |
| 3.3        | 电子文件归档               | 45        |
| 3.3.1      | 电子文件归档范围             | 45        |
| 3.3.2      | 归档收集方式               | 53        |
| 3.3.3      | 归档鉴定                 | 56        |
| 3.3.4      | 整理                   | 57        |
| 3.3.5      | 归档移交                 | 57        |
| 3.3.6      | 归档接收                 | 59        |
| 3.4        | 电子档案存储、保管、利用、处置      | 60        |
| 3.4.1      | 电子档案存储               | 60        |
| 3.4.2      | 电子档案保管               | 60        |
| 3.4.3      | 电子档案利用               | 65        |
| 3.4.4      | 电子档案处置               | 66        |
|            | 本章小结                 | 66        |
| <b>第4章</b> | <b>元数据</b>           | <b>67</b> |
| 4.1        | 元数据的基本概念             | 67        |
| 4.1.1      | 元数据的定义               | 67        |
| 4.1.2      | 元数据的概念内涵             | 69        |
| 4.1.3      | 元数据的分类               | 70        |
| 4.2        | 元数据在电子文件中的作用         | 73        |
| 4.2.1      | 电子文件管理需要元数据          | 73        |
| 4.2.2      | 元数据是电子文件管理的关键        | 75        |
| 4.3        | 元数据标准                | 76        |
| 4.3.1      | 国内外元数据管理标准研究         | 77        |
| 4.3.2      | 不同领域的元数据格式标准         | 77        |
| 4.4        | 电子文件元数据管理            | 85        |
| 4.4.1      | 元数据的管理过程             | 85        |
| 4.4.2      | 元数据的捕获节点             | 86        |
| 4.5        | 元数据项的详细定义            | 87        |
| 4.5.1      | 元数据项详细定义方法           | 87        |
| 4.5.2      | 元数据项详细定义实例           | 88        |
|            | 本章小结                 | 96        |

|                      |     |
|----------------------|-----|
| 第5章 电子文件知识库管理        | 97  |
| 5.1 电子文件管理与知识管理的融合   | 97  |
| 5.1.1 电子文件的知识属性      | 98  |
| 5.1.2 文件实体管理向知识管理的演变 | 100 |
| 5.2 知识与知识管理          | 104 |
| 5.2.1 数据、信息、知识和智慧    | 105 |
| 5.2.2 知识的定义和分类       | 108 |
| 5.2.3 知识管理           | 110 |
| 5.3 语义知识库技术          | 114 |
| 5.3.1 语义 Web         | 114 |
| 5.3.2 语义知识建模         | 119 |
| 5.3.3 语义知识推理         | 123 |
| 5.4 本体知识模型构建         | 125 |
| 5.4.1 本体描述语言         | 125 |
| 5.4.2 RDF/RDFS       | 127 |
| 5.4.3 OWL            | 133 |
| 5.4.4 本体建模工具 Protégé | 140 |
| 5.5 知识地图             | 150 |
| 5.5.1 知识地图的概念        | 150 |
| 5.5.2 知识地图的作用        | 152 |
| 5.5.3 知识地图的构建        | 153 |
| 本章小结                 | 154 |
| 第6章 电子文件知识分析         | 155 |
| 6.1 电子文件文献计量分析       | 155 |
| 6.1.1 文献计量经典定律       | 156 |
| 6.1.2 文献计量基本方法       | 159 |
| 6.1.3 文献评价计量指标       | 162 |
| 6.1.4 常用文献数据库        | 164 |
| 6.2 电子文件文本知识挖掘       | 167 |
| 6.2.1 文本挖掘的概念        | 168 |
| 6.2.2 文本预处理          | 169 |
| 6.2.3 文本挖掘算法         | 173 |
| 6.3 中文电子文件的分词        | 181 |
| 6.3.1 中文信息处理与分词      | 182 |
| 6.3.2 中文分词算法         | 183 |
| 6.3.3 中文分词系统         | 188 |
| 6.4 电子文件知识分析案例       | 191 |
| 6.4.1 案例背景           | 192 |



|                         |     |
|-------------------------|-----|
| 6.4.2 案例分析              | 192 |
| 本章小结                    | 195 |
| <b>第7章 电子文件网络信息资源管理</b> | 196 |
| 7.1 网络信息环境中的电子文件        | 196 |
| 7.1.1 网络信息资源的概念         | 197 |
| 7.1.2 Web 链接结构与信息内容     | 198 |
| 7.1.3 网络信息知识发现          | 203 |
| 7.2 网络信息获取              | 206 |
| 7.2.1 搜索引擎              | 206 |
| 7.2.2 网络爬虫              | 212 |
| 7.3 网络信息计量分析            | 217 |
| 7.3.1 网络链接分析            | 217 |
| 7.3.2 网络引文分析            | 226 |
| 7.3.3 社会网络分析            | 228 |
| 7.3.4 网络信息计量工具          | 236 |
| 7.4 网络信息计量分析应用案例        | 237 |
| 7.4.1 案例背景              | 237 |
| 7.4.2 案例分析              | 239 |
| 本章小结                    | 241 |
| <b>第8章 电子文件管理趋势与政策</b>  | 243 |
| 8.1 新技术环境下的电子文件管理       | 243 |
| 8.1.1 新技术条件下电子文件的安全管理   | 243 |
| 8.1.2 云计算环境下电子文件管理新趋势   | 246 |
| 8.1.3 电子合同与电子发票         | 247 |
| 8.2 电子文件管理体制            | 249 |
| 8.2.1 发达国家电子文件管理体制      | 249 |
| 8.2.2 我国的电子文件管理体制       | 255 |
| 8.3 电子文件管理政策法规          | 256 |
| 8.3.1 电子证据的法律保护         | 256 |
| 8.3.2 电子文件交换、传输和利用      | 257 |
| 8.3.3 电子文件版权保护          | 258 |
| 8.3.4 电子文件个人隐私保护        | 260 |
| 本章小结                    | 267 |
| <b>参考文献</b>             | 269 |

# 第 1 章

## 电子文件概述

与传统非数字形式的信息资源相比,数字信息资源表现出独特的性质。电子文件是信息技术的产物,信息技术的发展必然导致电子文件不断发生变化,形成新的特点。因此,文件和档案管理人员应当认真地、不断地研究电子文件的内涵、外延和特点,并据此探索电子文件科学管理的有效措施。在数字信息时代,电子文件具有法律证据效力。决定电子文件法律证据效力的不是其原始性,而是其真实性。

### 学习目标

本章以数字信息时代对于电子文件管理的影响为出发点,对电子文件管理进行概述。

- (1) 数字信息资源及其特性。
- (2) 数字信息时代的电子文件。
- (3) 文件、文档和档案。
- (4) 电子文件的定义和种类。
- (5) 电子文件与电子证据。

## 1.1 数字信息与电子文件管理

### 1.1.1 数字信息资源及其特性

在数字技术产生前,信息资源一直以非数字的形式存在,如纸质、模拟电磁信号等形式存在。但是,随着数字技术的发明和应用,信息资源开始以数字形式存在,并通过数字技术手段进行处理。这种以数字形式存在,并通过数字技术传输、处理、存储的信息资源统称为数字信息资源。

在国内诸多文献中,国内学者并未对数字信息资源进行定义,但是对信息资源概念的认识正在逐渐深化。有国内学者将信息资源定义为经过人类选取、组织、序化的有用的信息的集合。

与传统非数字形式的信息资源相比,数字信息资源表现出独特的性质,主要表现在以下方面。

#### 1. 信息的非人工识读性

从远古的泥板、石刻、羊皮纸、甲骨、竹简,到近代的纸质记录,人类世代代使

用的都是人工可以识读的记录符号。人可以在各种书写载体上直接写入或印刷各种信息，生成文件，也可以直接读出文件中的信息，并理解它的意思，不需要任何中介物。数字信息资源第一次使用了人工不可识读的记录符号——数字代码，人无法直接识读和理解经过复杂编码的“比特”，只有通过计算机特定的程序解码，使之还原为输入前的状态，人工才能识读它。

## 2. 系统依赖性

数字信息资源的制作、处理，以致存储的全部管理活动都必须借助于计算机系统才能实现，不兼容的计算机和应用软件生成的信息在交换使用时会遇到很大困难。当生成一份信息的软件、运行该软件的操作系统和硬件更新换代以致与原系统不兼容时，我们需要保存原系统，或者进行一系列适应新系统的转换、迁移工作，才能确保该份信息的可读性和可管理性。

## 3. 信息与特定载体之间的可分离性

自古以来所有信息资源记录，无论是文件还是图书，其信息都被固定在某一载体上，成为不可分离的一个“实体”，文献的使用、流传都是以这种实体形式实现的。对于正式文件来说，内容与最初形成时的标记同时被固定在这一实体上，使之打上了明显的原始性烙印。数字信息资源中的信息不再对原记录载体“从一而终”，不再具有物理意义上的固定实体状态，也不再具有固定的物理位置，可以在不同的载体上同时存在或相互转换，可以根据需要随时改变或扩展、缩小其存储空间。数字信息可以改变其在硬盘上的存址，可以从硬盘拷入软盘、磁带或光盘，可以从一张软盘、一盘磁带拷入其他软盘或磁带之中，还可以通过由电缆或光缆铺设的网络流向异地。数字信息资源中的信息是可流动的，具有相对独立性。

信息和特定载体的可分离性，还可以使数字信息资源以全文形式存在的“文件形态”与其中以逻辑关系连接的“信息要素”之间发生分离。随着关系模型的发展，调用若干个数据库中存储的数据“合成”数字信息资源将成为信息制作中经常使用的方法。这种“虚拟信息”的“合成命令”是程序，计算机执行一定的程序，可以将相关信息要素经逻辑组合形成一份信息，如果不将其全文保存下来，这份虚拟文件将不复存在。重复执行同样的指令，文件会迅速得到重组。但是，如果用于合成新文件的信息要素存在于动态数据库中，重组时数据库内信息可能已经发生变化，合成文件的内容也必然与前次有所不同。

## 4. 信息的易变性

造成数字信息资源信息发生变化的主要原因有如下三条：

(1) 在以往的信息资源中，载体不仅是信息的承载物，还具有一种对信息的固化作用，信息的原有顺序、本来面貌都被凝固在特定的载体上，不能改变。从前面所说的两种分离中我们看到，当信息可以脱离特定载体而存在时，载体对它的“束缚”就没有了，计算机系统中信息的相对独立性使人们对信息的增、删、更改十分方便，动态文档中的数据不断地被自动更新或补充。

(2) 数字信息资源载体性能的不稳定性有可能造成文件中信息的改变。到目前为止，数字信息资源存储介质的物理寿命（保证存储的数据不丢失的时间）还远远比不上纸张，

材料的氧化和变质、磁场的影响等很容易破坏磁介质上存储的数据,因此每隔一定时间就必须对原载体上数据进行拷贝,使之保持常新的状态,以防止信息的丢失。存储介质也在不断地更新换代。例如,软盘从8英寸到5英寸,再到3.5英寸,只用了不到10年时间,目前光盘已迅速普及起来。有理由预见,不久的将来还会有新的存储介质出现。随之更新的计算机逐渐对老式存储介质不再兼容,这就要求人们将原存储介质上的东西转换到新型存储介质上,以保证文件的可访问性,而转换过程中由于操作和其他方面的原因,可能导致信息的失真。

(3) 电子信息技术的发展,新的信息编码方案、存储格式、系统软件的不断出现更是对电子信息稳定的巨大冲击,因为新的计算机系统和程序常常无法解析那些过了时的代码,从而要求将原文件迁移到新的技术环境之中,迁移过程中信息的损失、变异也是不可避免的。

### 5. 信息存储的高密度性

电子文件的存储密度大大高于以往各种人工可识读的信息介质。随着技术的进步,电子文件介质的存储密度还将继续加大。

1956年9月,IBM公司(国际商业机器公司)制造的世界第一台磁盘存储系统只有区区的5MB。1995年个人计算机的硬盘容量达到了1.2GB(千兆字节),2000年达到20GB,2001年则猛升为40GB。与此同时,软盘、光盘的存储容量都在以很快的速度增长。从理论上讲,一个8GB硬盘约可存储40亿个汉字,一张3.5英寸高密度软盘(1.44MB)大约可存储70万个汉字,一张ZIP软盘(120MB~150MB)可存储6000万至7500万个汉字,一张5英寸光盘(650MB~750MB)可存储3亿~4亿个汉字或A4幅面的文稿图像数千页。应用信息压缩技术,光盘存储文字、图像文件的能力可增加数倍甚至更多,可谓“海量存储”。

### 6. 多种信息媒体的集成性

纸质文件主要承载文字和图形信息,而电子文件可以将文字、图形、图像、影像、声音等各种信息形式加以有机组合,这种文件称为“多媒体文件”。多媒体文件图、文、声、像并茂,能够更加真实地再现当时的活动情况,从而强化了文件对社会活动的记忆和再现功能。在用多媒体技术制作的文件中,文字、图像、声音等各种信息都用二进制数字来表达,它和文本文件一样可以在屏幕上显示,可以输出,并可以通过网络传输。

多媒体文件与其他声像文件的根本不同点在于数字信息和非数字信息的区别。以数字信息构成的各种电子文件不仅可以实现全文查询和网上远距离传输,用户还可以与之发生互动。例如,用户可以选择不同的输出方式(屏幕显示或打印、全文输出或局部输出、当地输出或异地输出、输出色彩与精度等),可以在“阅读”过程中调阅其他相关信息,在允许的情况下还可以对文件进行修改,从而使用户对于电子文件的利用乃至形成有了更大的参与性和选择权。

### 7. 信息的可操作性

电子文件中的数字信息不是静态、固态、消极的,而是动态、可变、积极的。数字信息的可操作性主要表现在两个方面。

1) 由于信息的易变性, 可以方便地改变其存在状态

在制作电子文件时, 制作者可以根据需要随心所欲地增、删、修改文件信息, 重新组织文件信息结构。如改变信息的原有位置, 剪切、复制信息, 多次地撤销或恢复对信息的操作。

(1) 可以调用存在于不同地方的数据组织文件, 在文件的任何地方加上表格和图形, 把文件制作成活动图像和声音。

(2) 用户甚至可以组织动态文件, 在适当的时候自动更新某些数据, 保持数据新颖、有效, 以及与其所反映事物状态的一致。

(3) 可以进行信息压缩, 提高存储密度和传输速度。这种压缩是一种可逆过程, 在解压之后, 信号可以完全恢复成原样。

(4) 可以改变其显示形式, 如将图形设计结果形成数字三维模型并进行动态旋转观看, 使设计人员看到其中的任何一个剖面及其有关的参数。

(5) 可以在纸质或胶片文件数字化转换过程中对其亮度、对比度及分辨率等进行调节, 以改善文件的清晰度, 对这些参数的局部调整对于工程图纸的局部消蓝, 以及档案蜕变字迹、图像的恢复具有很好的效果。

(6) 可以转换信息的媒体形式, 如将数字文本经扫描转换成数字图像, 经光学字符识别(OCR)又转换成数字文本; 将数字文本转换成数字语音, 或将数字语音转换成数字文本等。总之, 经过相应的操作, 人们可以使电子文件处于操作者希望的状态之下。

2) 可以利用已存在的文件信息做其他事情

电子文件中的信息不仅是自身含义的表达方式, 还可多方面地加以利用。

(1) 将原始信息输入相关的信息库, 形成各种资料库、文件库、零件库、部件库、材质库、色彩库、集成电路库等, 文件形成者制作电子文件时, 可使用查询软件通过关键字或指针等直接获取和交换相关库中的数据、图形、图像等信息, 形成多层次、立体化查询, 大大提高文件制作效率和质量。

(2) 在目前迅速发展的电子商务领域中, 以电子信息为基础, 通过软件和网络, 实现了无纸化的完整贸易过程。

(3) 在分析研究领域, 可以利用相关的软件, 对信息的模式进行识别, 对高空照片、X光片、各种图像进行识别, 对遥感遥测数据进行分析, 对自然语言加以理解和推理, 对数字化的文学作品进行研究, 然后自动作出判断, 形成结论。

(4) 在工程设计领域中, 经过整理的电子信息可以通过相应的软件对设计对象的各种性能进行仿真, 并根据给定数值对设计对象的强度、刚度等性能指标进行校核; 可以用CAD产品设计文件中的三维几何模型模拟零件、部件和设备的装配与安装; 可形成仿真模型或随机生成各种模拟环境, 用于实验和操作人员的模拟训练。如波音公司设计777喷气式飞机时, 用一个包容了所有零部件以及它们在一起是如何工作的数字三维模型代替了过去的蓝图和实物模型, 节省了近10万张图纸以及绘图和复制所需的大量人力和时间, 也节省了昂贵的飞机全尺寸模型。

(5) 在生产实践领域中, 特定的电子信息可以直接进入NC系统(数控车床)或CAM系统(计算机辅助制造)控制生产, 进行产品加工和制造。在管理领域, 与相关的管理系统、工作系统相链接, 可实现自动业务管理和事务管理。如系统根据给定

的要求自动进行各种事务提示；将 CAD 产品设计系统与材料价格单、库存材料清单等组成统一的设计平台，将电子文件的制作、管理与成本管理、库存管理自动结合起来等。

(6) 计算机智能研究领域更是以电子信息为基础，涉及了诸如数学、心理学、生物学、语言学、逻辑学、经济学、哲学、法学、医学、计算机科学等许多重要学科。构建的专家系统和具有学习功能的神经网络系统、模糊控制系统广泛地应用在各个领域，从而代替了人类的某些脑力劳动。

因此，我们已经掌握和尚未掌握的电子信息的种种可操作性能，赋予电子文件以活的生命和更广泛的用途。

掌握电子文件的特点是科学地管理电子文件的前提。电子文件是信息技术的产物，信息技术的发展必然导致电子文件不断发生变化，形成新的特点。因此，文件和档案管理人员应当认真地、不断地研究电子文件的内涵、外延和特点，并据此探索电子文件科学管理的有效措施。

## 1.1.2 数字信息时代的电子文件

### 1. 电子文件已经成为经济社会的基本信息支撑

自 20 世纪 90 年代以来，在我国政府高层领导对信息化发展长期、持续的关注下，在各级政府的有力领导下，我国已经确定了信息化在经济社会发展中的主导地位。1993 年，首批“金”字工程启动。1997 年，《国家信息化“九五”规划和 2010 年远景目标》构建了信息化建设的框架。2000 年，中国共产党第十五届中央委员会第五次全体会议通过《中共中央关于制定国民经济和社会发展第十个五年计划的建议》，指出：“大力推进国民经济和社会信息化，是覆盖现代化建设全局的战略举措。”2001 年，国家信息化领导小组成立，同年推进电子政务建设的指导方针出台。2006 年 5 月，《2006—2020 年国家信息化发展战略》发布，标志着信息化进入一个崭新的阶段。

在连续的强力政策推动之下，信息技术应用迅速向各行各业渗透，组织机构正在由基于纸张的工作环境向电子环境迁移，许多领域正逐步实现业务活动电子化、无纸化。信息系统覆盖面和应用速度就是电子文件的产生面和产生的速度，结果是惊人的。以中央机构为例，2007 年度中国科协重点课题《我国电子文件管理机制研究》研究小组针对 49 家中央和国家机关及其直属企事业单位的调查表明，中央和国家机关及其直属企事业单位电子文件数量占全部文件数量的 72.7%。以北京市政府为例，到 2008 年底，北京市 1920 项行政办事事项中，100% 提供了网上办事指南服务，其中 1100 多项实现网上申报、状态查询和结果公示。截至 2009 年 6 月，北京市政府建成了 600 多个信息系统，覆盖宏观决策、经济调控、市场监管、公共服务、城市管理等领域，对政府核心业务覆盖率达 90% 以上。再以企业为例，2007 年，我国 1.5 万个国有大中型企业中，80%~90% 的企业建立了不同规模的计算机辅助管理系统，包括 MIS 和 OA 系统等；几千家引进 MRP、ERP 软件，约 1/3 取得了较好效果。机械、电子、航空、造船、石油、化工等行业中，80%~90% 的大中型企业主要产品采用 CAD 设计，主要生产过程采用 CNC、PLC、DCS 进行控制。由此得到一个基本判断，我国大中型企业信息技术应用基本普及，大部分已

实现中级信息化,部分大中型企业集团已进入高级信息化阶段。2006年,在研发过程中采用了任何一项数字化设计技术的制造企业中,使用二维CAD设计技术的企业已经达到了90%,使用三维CAD设计技术的企业已经达到了82%。

## 2. 电子文件普遍应用

赛迪研究院指出,2006—2007年,我国整个计算机市场的明星用户已从大中型企业变为中小型企业,从“金”字工程的行业转为公共服务部门,竞争的重点由中心城市向地方转移。这说明信息化已经全面铺开,不但电子文件的数量激增,其应用范围也日益扩大,电子文件的应用不再是信息化水平高的地区、行业、单位的特有现象,而是几乎所有地方、行业 and 单位面临的共同趋势。从电子投票到网络会议,从无纸制造到网上商城,从数字影视到博客、维基,电子文件的应用已经渗透到政治、经济、文化等社会生活的方方面面。

国际化程度的加深进一步推动了我国信息化的进程,也加快了电子文件的全面应用。比如,根据中国民航信息网络股份有限公司与国际航空运输协会(IATA)的协议,到2007年底,国内所有机场采用电子客票。许多涉外工程、外包项目中,建设方和业主方交接的资料也以电子方式进行。

在电子文件应用普及的背景下,电子文件的种类日益丰富,包括字处理文件、数据库文件、图形文件、音频文件、视频文件、多媒体文件、超媒体文件等多种类型,电子文件格式多样化、更新快成为其直接后果。据估算,全世界有2万~3万种文件格式。仅在办公领域,目前世界上就有1万多种办公文件格式,常用文档格式有300多种。各种文件制作软件每隔3~5年就有更新,也会引起格式的变化。比如在办公领域中市场占有率最高的Microsoft Office,10年间就拥有了Office 97、Office 2000、Office XP、Office 2003、Office 2007这五个版本,Office 2010也在2010年发布。

## 3. 电子文件增量可观

由于信息技术的应用极大地扩展了人们产生文件的能力,因此理论上讲电子文件的数量将不断增长。对比信息化水平较高的国家的统计数据,就可以看出我国电子文件数量的发展趋势。截至2007年,美国国家档案馆已经保存了包括国会、法院、总统办公室、总统委员会和将近100个办公署、部门、分支机构及其承包商的将近100亿份电子文件。而国家档案局进行的调查显示,已有近80%的中央和国家机关、中央企业采用办公自动化或电子政务系统,产生各类电子文件近2亿份。也就是说,我国中央机关文件形成的数量仅为美国联邦机关移交进馆数量的2%,即便是统计口径存在一些偏差,其差距也非常惊人。

纵向比较来看,“十一五”之后,我国进入信息化快速推进阶段。在这个阶段,电子文件的增幅异常明显。以北京市为例,北京市2008年建成的政府信息系统对政府核心业务覆盖率达85%,比2007年提高20%。2006年,一项对企业信息化的调查显示,在研发过程中采用了任何一项数字化设计技术的制造企业中,使用三维CAD设计技术的企业已经达到了82%,而仅仅一年前,三维CAD普及率只有63%。上海市2007年电子商务的总额度是2002年的10倍,占全市商品贸易额度的比例由2.78%提高到11.87%,增长了三倍多。

横向比较来看,鉴于中国信息化的快速进程,中国的信息增长速度要高于全球。IDC 研究报告《数字宇宙膨胀:到2010年全球信息增长预测》指出,2006年全球每年制造、复制的数字信息量共计1610亿GB,中国数字信息量为127.1亿GB,占全球信息量的7.9%;2012年,IDC在《数字世界研究》中提出,中国在全球数字世界中所占的份额将从2012年的13%增至21%。这个结果表明我国电子文件的增长速度在全球也靠前。

美国《纽约时报》2008年11月27日报道,布什政府电子数据储量大约为100TB,这一数值约为前总统比尔·克林顿两届政府档案总量的50倍,是国会图书馆2000万册编目图书量的5倍。在我国电子文件增速高于全球平均速度的背景下,美国的数据为我们描摹了中国电子文件的增幅发展趋势,这样的增幅可用指数级来形容。

## 1.2 电子文件的概念

### 1.2.1 文件、文档和档案

#### 1. 文件

文件是国家机构、社会组织或个人在履行其法定职责或处理事务中形成的各种形式的信息记录。

文件不只是一个数据的集合,还是一个事件的结果或产物,它与业务活动相关联。文件的一个显著特点就是其内容必须以固化形式存在,即业务事项处理固化的表现形式,它具有以下特点。

(1) 文件形成、接收和使用的主体是各种社会组织或个人。

(2) 文件具有最直接的应用性。文件形成、接收和使用的目的是满足各类社会组织或个人记载事物、表达意志、交流信息、联系工作、处理各种事务的需要,具有最直接的应用性,是处理事务的必不可少的一种工具。

(3) 文件具有一定的规范体式,包括法定和约定俗成的体式。文件体式具有长期的稳定性和固定表达方式,以维护文件的严肃性、郑重性和权威性。

(4) 凭证属性和证据价值。文件伴随着各项管理活动同步产生,是对该项活动的原始记录,对一定的组织或人员具有法律所规定的约束力和影响力,在法律上具有不可否认的凭证属性。因而文件不仅具有信息价值,还具有证据价值(evidential value)。

(5) 必须包括内容、背景和结构三种基本要素,以保证其凭证价值。

(6) 文件的形式多样,其形式或载体不影响其作为文件的用途和价值。

#### 2. 文档

文档是为防止信息丢失或遗忘,将其记录并保存在实体介质上的信息,是固定在载体上的信息或数据。文档包括内容和元数据,可以作为一个整体在应用背景下读出。因此,文档是指可视为一个整体对待的记录信息或对象。

从两者的定义上看,文档与文件既互相联系又有不同。两者是相交关系,都是信息记录。文档与文件的不同之处在于,文档可以被编辑和删除,而文件必须采取控制措施使其处于固定的状态。



图 1-1 描述了文件与文档的关系。

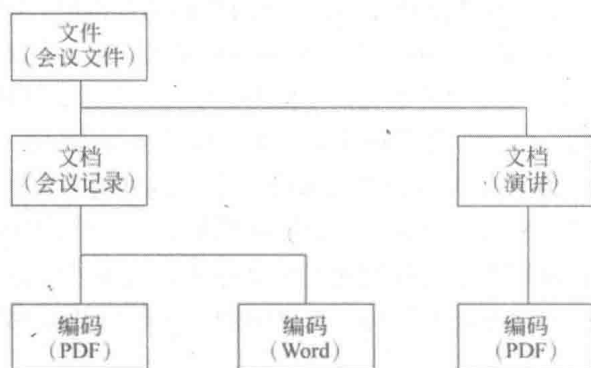


图 1-1 文件与文档的关系

### 3. 档案

根据《中华人民共和国档案法》定义，档案是指是过去和现在的国家机构、社会组织以及个人从事政治、军事、经济、科学、技术、文化、宗教等活动直接形成的对国家和社会有保存价值的各种文字、图表、声像等不同形式的历史记录。

档案的这个定义具有如下含义。

(1) 说明了档案的形成者。档案是由官方机构、半官方机构、非官方机构以及一定的个人、家庭和家族形成的。

(2) 说明了档案来源于文件。档案是由文件有条件地转化而来的，这里的“文件”是指广义文件，即一切由文字、图表、声像等形式形成的各种材料。档案和文件是同一事物在不同价值阶段的不同形态，两者具有同源性和阶段性的共性，也具有实效、功用、离合等个性差异。从文件到档案是一个批判继承的辩证运动过程。从信息的内容和形式来说，两者是完全相同的，但从时效、价值和系统性上来说，档案是对文件的不断扬弃。首先是时效性批判，档案是已经办理完毕的文件；其次是价值评判，档案是办理完毕的文件中具有保存价值的部分；最后是系统性批判，档案是把分散状态的文件按一定逻辑规律整理而成的信息单元。因此，文件是档案的前身，档案是文件的归宿；文件是档案的基础，档案是文件的精华；文件是档案的素材，档案是文件的组合。

(3) 说明了档案的形式多种多样。档案的形式包括载体、制作手段、表现方式等。从载体来看，有甲骨、金石、缣帛、简册、纸质、胶磁等；从制作手段来看，有刀刻、笔写、印刷、复制、摄影、录音、摄像等；从表现方式来看，有文字、图表、声像等。

(4) 指出了档案的本质属性。档案具有历史再现性、知识性、信息性、政治性、文化性、社会性、教育性、价值性等特点，其中历史再现性为其本质属性，其他特点为其一般属性。

## 1.2.2 电子文件定义和种类

### 1. 电子文件的概念

电子文件 (Electronic Records) 是以代码形式记录于磁带、磁盘、光盘等载体，依赖计算机系统存取并可在通信网络上传输的文件。