

大数据环境下 电子文件管理元数据 研究与实践

■ 王大青 张新建 蒙泓 著

图书在版编目 (CIP) 数据

大数据环境下电子文件管理元数据研究与实践 / 王大青等著.
—成都: 四川人民出版社, 2015. 8
ISBN 978-7-220-09559-7

I. ①大… II. ①王… III. ①电子档案—档案管理
—研究 IV. ①G276

中国版本图书馆 CIP 数据核字 (2015) 第 174114 号

DASHUJU HUANJINGXIA DIANZI WENJIAN GUANLI YUANSHUJU
YANJIU YU SHIJIAN

大数据环境下电子文件管理元数据研究与实践

王大青 张新建 蒙泓 著

责任编辑
封面设计
版式设计
责任校对
责任印制

王 茜
解建华
戴雨虹
蓝 海
王 俊

出版发行
网 址
E-mail
新浪微博
发行部业务电话
防盗版举报电话
照 排
印 刷
成品尺寸
印 张
字 数
版 次
印 次
书 号
定 价

四川人民出版社 (成都槐树街 2 号)
<http://www.scpph.com>
scrmcbs@sina.com
@四川人民出版社官博
(028) 86259624 86259453
(028) 86259624
四川胜翔数码印务设计有限公司
四川福润印务有限责任公司
160mm×225mm
22.25
310 千
2015 年 8 月第 1 版
2015 年 8 月第 1 次
ISBN 978-7-220-09559-7
32.00 元

■ 版权所有·侵权必究

本书若出现印装质量问题, 请与我社发行部联系调换
电话: (028) 86259453

前 言

在人类由信息时代迈向知识时代的进程中，电子文件管理实践向电子文件管理元数据提出了理论发展与创新的要求。从大背景来看，知识时代、大数据环境要求开展基于本体的电子文件管理元数据集成研究；从具体实践看，电子政务建设、云计算服务模式、本体和关联数据的应用等社会实践均要求对基于本体的电子文件管理元数据理论与实践进行研究。

本书在充分调研国内外电子文件管理元数据前沿进展的基础上，揭示了电子文件的知识资产属性和价值运动规律，深刻地分析了发展中的文件和档案学理论对电子文件管理元数据立体、多维的支撑作用。从电子文件相关概念出发，系统地论述和丰富了电子文件管理元数据的基本理论，揭示了电子文件的知识属性。而后，从时代发展趋势出发，创新地研究了知识时代、大数据环境、云计算服务模式、本体和关联数据应用等对电子文件管理元数据的拓展需求，要求档案工作者从知识管理的视角出发、科学地管理电子文件，即实施基于本体的电子文件管理元数据集成。

本书在电子文件管理元数据的应用研究中，将本体作为知识表示的重要方法。在系统论述电子文件管理元数据集成理论的基础之上，本书紧扣电子文件前端控制的关键——电子文件分类方案，也就是电子文件分类元数据控制体系，开展了基于集成的电子文件分类方案研究，即为



了电子文件分类方案应用与维护的需要，将电子文件分类方案、保管期限表与处置表集成起来，详细论述了其框架设计、编制原则、实现路径和具体元数据构成，并创建了本体实例。

随后，本书设计了电子文件管理元数据应用架构和集成应用的总体模型。将电子文件管理元数据国际标准 ISO 23081 的高层元素与具有全程控制作用的档案著录系列国际标准中的元数据进行了集成，并开展了基于本体的集成应用研究，具体地构建了本体应用的多维模型。包括与国际档案著录规则（总则）、规范记录标准和职能著录标准等档案元数据的集成应用。这些无论是对于电子文件管理元数据理论的发展与丰富，还是对于我国电子文件管理元数据标准的完善和电子文件管理元数据方案的具体制定，并在实践中充分发挥元数据的组织与控制功能、建立起电子文件知识组织体系，均有着十分重要的意义。

本书研究的总体方向以及立足点为面向知识时代的电子文件管理元数据，研究的聚焦点是大数据环境下知识组织对电子文件管理元数据所提出的具体要求，两者具有呼应关系。在具体实践研究中，本书以知识组织这条主线贯穿于始终。

因作者水平有限，本书错漏之处在所难免，敬请广大读者批评指正。

王大青 张新建 蒙 泓

二〇一五年八月

目 录

00 绪 论

0.1 研究背景

0.1.1 时代背景

0.1.2 社会背景

0.1.3 技术背景

0.1.4 专业背景

0.2 研究目的与意义

0.2.1 研究目的

0.2.2 研究意义

0.3 主要研究内容

01 国内外研究现状及其述评

1.1 国外研究现状

1.1.1 国际标准及国际协作项目研究现状

1.1.2 各国研究现状

1.2 国内研究现状

1.2.1 从国家标准层面上看电子文件管理元数据研究进展

1.2.2 从理论发展视角看电子文件管理元数据研究进展

1.2.3 从实践应用视角看电子文件管理元数据研究进展

1.3 国内外研究现状述评



1.3.1 国外研究现状述评

1.3.2 国内研究现状述评

02 电子文件管理元数据基本理论及其发展

2.1 对电子文件的再认识

2.1.1 电子文件概念定义的深化与拓展

2.1.2 基于文件和档案管理视角的电子文件相关概念

2.1.3 基于计算机应用视角的电子文件相关概念

2.1.4 电子文件的知识资产属性

2.1.5 电子文件价值运动规律

2.2 电子文件管理元数据概念

2.2.1 文件管理元数据概念

2.2.2 电子文件管理元数据概念

2.3 发展的文件和档案学理论对电子文件管理元数据的支撑

2.3.1 新来源观及其对电子文件管理元数据的支撑

2.3.2 后保管模式及其对电子文件管理元数据的支撑

2.3.3 文件连续体理论及其对电子文件管理元数据的支撑

2.3.4 开放档案信息系统参考模型及其对电子文件管理元数据的支撑

2.3.5 电子文件管理原则的发展

2.4 电子文件管理元数据类型

2.4.1 基于电子文件运动阶段划分的电子文件管理元数据类型

2.4.2 基于实体划分的电子文件管理元数据类型

2.4.3 基于属性划分的电子文件管理元数据类型

2.4.4 基于系统划分的电子文件管理元数据类型

2.5 电子文件管理元数据的作用与功能

2.5.1 电子文件管理元数据作用

2.5.2 电子文件管理元数据的具体功能

2.6 电子文件管理元数据的基本结构

2.6.1 电子文件管理元数据的宏观结构

2.6.2 电子文件管理元数据的微观结构

本章小结

03 电子文件管理元数据需求的时代拓展

3.1 电子文件管理元数据需求分析框架

3.1.1 电子文件管理元数据需求分析概念

3.1.2 电子文件管理元数据需求分析的意义

3.1.3 电子文件管理元数据需求分析目标

3.1.4 电子文件管理元数据需求分析原则

3.1.5 电子文件管理元数据需求分析方法

3.1.6 电子文件管理元数据需求分析流程

3.2 电子文件管理元数据需求的拓展

3.2.1 电子文件管理元数据基本需求

3.2.2 知识服务驱动的电子文件管理元数据需求

3.2.3 电子政务建设驱动的电子文件管理元数据需求

3.2.4 云计算服务模式驱动的电子文件管理元数据需求

3.2.5 大数据环境驱动的电子文件管理元数据需求

3.2.6 本体应用驱动的电子文件管理元数据需求

3.2.7 关联数据应用驱动的电子文件管理元数据需求

本章小结

04 电子文件管理元数据的集成

4.1 电子文件管理元数据集成概念及意义

4.1.1 集成理论及其应用

4.1.2 电子文件管理元数据集成基本内涵及其优势

4.2 电子文件管理元数据集成框架

4.2.1 电子文件管理元数据集成理论模型

4.2.2 电子文件管理元数据集成实施架构



4.2.3 电子文件管理元数据交换与映射模型

4.3 电子文件各类实体元数据的集成

4.3.1 电子文件实体元数据的集成

4.3.2 电子文件责任者实体元数据的集成

4.3.3 电子文件业务实体元数据的集成

4.3.4 电子文件关系实体元数据的集成

4.3.5 电子文件、责任者、业务与关系实体元数据的综合集成

本章小结

05 基于集成的电子文件分类方案理论及其应用

5.1 电子文件分类方案概念与作用

5.1.1 电子文件分类方案概念

5.1.2 电子文件分类方案作用

5.2 国内外电子文件分类方案需求

5.2.1 国际标准 ISO 15489 的文件分类方案需求

5.2.2 国际标准 ISO 16175 的文件分类方案需求

5.2.3 美国国防部 DoD5015.02-STD 的文件分类方案需求

5.2.4 欧盟 MoReq2010 的文件分类方案需求

5.2.5 我国电子文件管理标准中的文件分类方案需求

5.3 电子文件分类方案所应用的基本分类方法

5.3.1 基于职能的分类方法

5.3.2 基于职能分类的原因

5.3.3 基于职能分类的意义

5.4 基于集成的电子文件分类方案基本理论

5.4.1 电子文件分类方案实施集成管理的必要性

5.4.2 基于集成的电子文件分类方案具体功能

5.4.3 基于集成的电子文件分类方案需求分析

- 5.4.4 基于集成的电子文件分类方案基本结构
- 5.4.5 基于集成的电子文件分类方案构建原则
- 5.4.6 基于集成的电子文件分类方案构建路线
- 5.5 基于集成的电子文件分类方案本体应用实践
 - 5.5.1 基于集成的电子文件分类方案语义应用
 - 5.5.2 基于集成的电子文件分类方案本体构建实例
- 本章小结
- 06 基于本体的电子文件管理元数据语义集成应用
 - 6.1 电子文件管理元数据应用架构
 - 6.1.1 电子文件管理元数据应用总体框架
 - 6.1.2 电子文件管理元数据 Schema 的制定与应用原则
 - 6.1.3 电子文件管理元数据应用总体要求
 - 6.1.4 电子文件管理元数据应用流程
 - 6.2 电子文件管理元数据集成应用总体模型
 - 6.3 基于 ISO 23081 与 ISAD (G) 的电子文件管理元数据集成应用
 - 6.3.1 电子文件特征说明项
 - 6.3.2 电子文件背景信息项
 - 6.3.3 电子文件内容与结构项
 - 6.3.4 电子文件检索与利用项
 - 6.3.5 电子文件关系项
 - 6.3.6 电子文件著录控制项
 - 6.3.7 电子文件附注项
 - 6.4 基于 ISO 23081 与 ISAAR (CPF) 的电子文件管理元数据集成应用
 - 6.4.1 责任者特征说明项
 - 6.4.2 责任者描述项
 - 6.4.3 责任者关系项



- 6.4.4 规范记录著录控制项
- 6.5 基于 ISO 23081 与 ISDF 的电子文件管理元数据集成应用
 - 6.5.1 职能活动特征说明项
 - 6.5.2 职能活动背景项
 - 6.5.3 职能活动关系项
 - 6.5.4 职能活动著录控制项
- 6.6 基于本体的电子文件管理元数据集成应用语义分析
 - 6.6.1 电子文件管理元数据集成应用本体构建目标
 - 6.6.2 本体概念体系构建原则
 - 6.6.3 电子文件管理元数据集成应用本体核心概念集
 - 6.6.4 电子文件管理元数据集成应用本体概念关系分析
 - 6.6.5 电子文件管理元数据集成应用本体逻辑模型
 - 6.6.6 基于本体的电子文件管理元数据集成应用关联分析
 - 6.6.7 基于本体的电子文件管理元数据集成应用实例
- 6.7 不同类型客观对象本体描述逻辑架构
 - 6.7.1 电子文件管理领域本体逻辑框架
 - 6.7.2 人物本体
 - 6.7.3 组织机构本体
 - 6.7.4 地理本体
 - 6.7.5 事件本体
 - 6.7.6 法规本体
 - 6.7.7 电子文件本体
 - 6.7.8 时间本体

本章小结

07 基于 OWL2 的电子文件管理元数据语法集成应用

7.1 基于 OWL2 的电子文件管理元数据集成应用本体形式化描述

- 7.1.1 基于 OWL2 的本体结构
- 7.1.2 OWL2 语法及其在电子文件管理元数据集成中的应用
- 7.1.3 OWL2 推理机制及其在电子文件管理元数据集成中的应用
- 7.1.4 基于 OWL2 的电子文件管理元数据集成应用本体形式化描述的实现
- 7.2 基于本体的电子文件管理元数据集成应用系统
 - 7.2.1 基于本体的电子文件管理元数据集成应用系统需求分析
 - 7.2.2 基于本体的电子文件管理元数据集成应用系统总体架构
 - 7.2.3 基于本体的电子文件管理元数据集成应用系统实现
 - 7.2.4 基于本体的电子文件管理元数据集成应用系统原型

本章小结

08 总结与展望

附录 基于 OWL2 的电子文件管理元数据代码示例

参考文献

00

绪 论

时间与空间是物质运动的基本属性。任何物质的运动都会以一定的时间和空间方式、特性表现出来。

——马克思

在人类由信息时代迈向知识时代的进程中，电子政务、电子军务和数字政府建设的推广，电子商务的发展，智慧城市建设的兴起，物联网的大规模应用，智能手机的普及等等共同加快了数据的生产速度，人类进入了一个前所未有的大数据环境。为此，2012年召开的第17届国际档案大会，以“变化的环境”为中心主题，探讨档案工作者所面临的挑战。在变幻的时代背景中，知识服务、大数据环境、电子政务和云计算等，对传统档案理论和实践模式形成了重大冲击，对档案理论进行重构、对档案管理实践进行创新，是时代赋予我们的重要职责。

今天，人类社会以前所未有的速度向前发展，其中一个重要原因就是知识的网络化共享与应用，以及由此而带来的更为高效的知识创新。电子文件，不仅是知识的重要源泉，更是当下档案的核心构成部分。电子文件，真实地记录了组织机构和个人的活动，可以系统地反映人类社会历史面貌，是社会记忆不可或缺的构成部分，具有重要的情报价值、凭证价值和文化价值。许多电子文件不仅仅对当今时代具有重要作用，而且在人类社会发展的历史长河中具有举足轻重的长久保存价值。科学



地管理电子文件是全人类的共同职责。时代的发展、社会的进步、技术的革新、实践的需求，强烈要求我们从战略视角，以前瞻和发展的眼光审视电子文件管理元数据理论与实践。

从理论上讲，在新的时代背景之下，电子文件的管理需求，对传统载体档案管理理论提出了重要挑战。在传统理论体系被打破、新的管理理论体系建立与完善之际，电子文件管理理论的系统性、前瞻性和实践指导性有待进一步增强，尤其是电子文件管理的“命脉”——电子文件管理元数据的理论研究，起步较晚，更是不可避免地存在着一些亟待解决的问题。

从实践上看，各级都在一定程度上开展了电子文件管理工作。开始注重电子文件管理国家战略的制定，近年来先后出台了电子文件管理专项国家标准和行业标准，安徽、江苏、上海、浙江、江西、福建和天津等省市先后开展了具有代表性的电子文件管理实践探索。但是，从总体效果来看，国家对电子文件这种战略性资源的掌控能力依然严重不足^①，电子文件管理元数据对电子文件的全程控制功能没有得到应有的重视与发挥，仍然存在着电子文件管理元数据标准体系不完善、管理理念滞后、应用性较强的元数据研究成果不足等问题，由此带来了电子文件流失严重、管理规范化程度不高、低水平重复建设普遍等现象。在电子文件利用中，存在着检索效率不高、以简单信息服务为主、大量有价值电子文件得不到充分利用等问题。

面向知识时代，密切关注大数据环境下知识服务的发展趋势，充分利用本体、云计算、关联数据等信息技术和模式，超前思考电子文件管理工作的未来图景，是档案领域科研的重要内容。对于中国文件和档案领域而言，扭住电子文件管理的根本——电子文件管理元数据，开展其理论与应用研究显得尤为迫切。这项研究的宗旨就是要从电子文件

^① 冯惠玲，刘越男. 电子文件管理国家战略 [M]. 北京：中国人民大学出版社，2011，第27页.

的设计、形成开始，直至电子文件进入永久性保管机构，建设科学、规范的电子文件管理元数据体系，应用本体方法提升电子文件管理的智能化，这对于有效应对当今时代的电子文件管理挑战，确保电子文件的真实性、可靠性、完整性和可用性具有重要意义。尤其是对于大数据环境下，我国政府、军队信息与知识的集成管理与集成服务具有直接指导意义。

0.1 研究背景

0.1.1 时代背景

马克思主义唯物辩证法告诉我们：一切事物的运动规律是与时间、空间息息相关的。因而，我们进行电子文件管理元数据研究时，就不能不从时间轴上考察一下我们所处的时代。也就是说，只有认清了时代背景，才可能站在全局高度对“电子文件管理元数据”有一个科学的认识。

联合国教科文组织于2005年向全世界正式发布的《迈向知识社会》的时代报告给出了权威的说明，在这个报告中专门指出^①：

随着国际互联网新技术的不断出现与日益发展，知识已成为社会发展的源泉，人类社会正在由信息社会迈向知识社会。所谓的知识社会就是具有对信息进行确认、生成、处理、转换、传播及使用并为人类的发展构建知识与应用知识的社会。

该报告明确指出人类社会正由信息社会迈向知识社会。

早在1998年，世界银行就发布了《知识促进发展》的报告，指出：在人类社会的发展中，以知识为动力的经济发展战略正在形成^②。知识

^① United Nations Educational, Scientific and Cultural Organization. Towards Knowledge Societies [R]. France: Paris, 2005.

^② The World Bank. Knowledge for Development [R]. (1999). [2013-10-31]. <http://www.worldbank.org/wdr/>.



的生产、传播和应用成为生产力发展的重要驱动力，在全世界范围内出现了基于知识的产业链。信息流、知识流在世界范围内的快速流转和共享，加快了时代发展的步伐。信息和通信技术的发展为知识社会的来临创造了条件。在2012年联合国教科文组织的年度报告中，再次将“建设知识社会”作为专门一章进行重点论述^①。

由此可见，当今我们所处的时代正由信息时代迈向知识时代。这是由知识在人类社会组织、管理和发展等方面的诸多优越性及其作用所决定的，是时代发展的必然趋势。著名的未来学家托夫勒在《第三次浪潮》中指出，“随着信息与经济的全球化、网络化，知识将成为最为关键的社会经济资源”^②。

在当今时代的发展进程中，知识驱动着社会改革、知识促进着社会发展、知识创造着社会财富。知识，正以其多元的存在形式，成为当代社会的关键资产^③，并快速地在社会各个领域、各个层面以知识管理取代信息管理，以知识服务取代信息服务，从而驱动着人类社会发生深刻的变化。为此，著名的美国经济学家和管理学专家彼得·德鲁克认为：在知识时代，最基础的经济资源是知识，知识工作者将在社会发展中发挥核心作用^④。著名的国际档案学专家加拿大学者特里·库克早在1994年就指出：“档案工作者不能囿于档案保管员的角色，而应成为理念与知识的提供者，将着眼点从信息转移至知识上来”^⑤。

① UNESCO. UNESCO 2012 [R]. (2013-03). [2013-12-15]. <http://unesco.org/images/0022/002204/220416e.pdf>.

② Alvin Toffler. [The Third Wave](#) [M]. USA Bantam Books, 1980.

③ Roberta Cuel. A knowledge-oriented analysis of organizations. Knowledge Nodes: the building blocks of a Distributed Knowledge Management Approach [EB/OL]. [2013-10-31]. <http://fandango.cs.unitn.it/~rcuel/docs/dceis-cuel.pdf>.

④ Drucker Peter F. [Management Challenges of the 21st Century](#) [M]. New York: Harper Business, 1999.

⑤ Terry Cook. Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era [J]. [Archives and Manuscripts](#) 22, 1994 (2): 300-328.

电子文件，作为知识资产，积淀了人类在征服自然、改造自然过程中的智慧，忠实地记录了人类在社会生产和社会生活中的行为，反映了人类的认知水平，是人类智慧的结晶。当今时代，电子文件在档案中所占的比重呈直线上升趋势。可以这样说，在人类由信息时代迈向知识时代的进程中，电子文件不可或缺、无可替代。我们不但要保护好这一珍贵资源，而且要按照知识服务的要求高效、高水平地提供电子文件利用，这也是信息服务领域一个共同的时代性课题。因此，我们要抓住电子文件的本质属性——知识资产属性，站在时代至高点，以科学的、历史的、前瞻的眼光思谋电子文件管理元数据的创新与发展，从而推动电子文件管理模式与时俱进地创新和发展。

0.1.2 社会背景

对“电子文件管理元数据研究”所处的社会背景进行考察，是要对影响课题研究的诸多社会要素进行的详细分析。唯有如此，才可能对本书有一个全面、多维的深入认识。

社会学创始人孔德认为：社会作为由各个不同要素构成的整体系统，社会结构各个部分之间的平衡与和谐关系，是社会正常运转的基本条件。当今时代，作为社会正常运转的各要素之特征的总和就构成了社会的总体特征——当代社会背景。从影响电子文件管理的关键因素看，当今社会背景主要特征有：网络化社会、数字政府、智慧城市和知识经济。

0.1.2.1 网络化社会

所谓网络化社会^①，就是指电子网络构建了新社会形态，它改变了生产、经济、政治与文化发展的过程和结果，文字、人际关系、数据和财富这些社会要素都可以在网络空间中得到体现。随着信息和网络技术的发展，尤其是互联网在世界范围内的普及（截至2013年12月底，我

^① 冯务中，李艳艳，“网络社会”概念辨析 [J]. 广西社会科学，2008（9）：175.



国网民 6.18 亿人，互联网普及率达 45.8%，手机网民 5 亿^①；全世界网民达 27 亿^②），人类社会也在网络驱动下发生着悄然变革。美国“9·11”数字档案馆的建立与埃及“颜色革命”网络电子文件收集工程以无可辩驳的事实向世人诠释了网络化社会中电子文件的重要价值。

0.1.2.2 数字政府

数字政府，又称为虚拟政府或电子政府^③，其建设并非以计算机和网络技术简单地替代传统的政府运作模式，而是应用信息、网络和通信等技术，对政府机构的业务运行、管理模式和服务模式乃至服务理念进行优化，实现政府运作和服务模式的创新发展，以提高政府工作效率和服务水平。在数字政府建设中，美国率先将数字政府列入政府改革议程。1993 年克林顿政府建立了“国家绩效评估委员会”，提出了“电子政府”概念。2012 年 5 月，美国联邦政府又专门发布了数字政府建设战略——《数字政府：构建 21 世纪平台》^④。当前，数字政府建设已成为世界各国关注的焦点。数字政府与传统运作模式的政府相比，运作效能显著提升，由此，其成为一个国家综合国力的重要衡量指标。在数字政府建设中，关键的一条就是知识管理和应用。随着数字政府的高效运行，电子文件数量迅速增长，如何对其进行有效控制、组织、开发和利用就成为当今数字政府建设一个迫切需要解决的问题。

0.1.2.3 智慧城市

以数字政府为基础，智慧城市也出现在人类社会的发展之中。2006

① 中国互联网络信息中心. 第 33 次中国互联网络发展状况统计报告 [EB/OL]. (2014-01). [2014-05-20]. <http://www.cnnic.net.cn/hlwfzyj/hlwfzxx/qwfb/201403/w02014032736969016832.pdf>.

② 环球网科技. 联合国：手机用户数量 2014 年将超过全世界人口 [EB/OL]. (2013). [2013-11-08]. <http://tech.huanqiu.com/comm./2013-05/3929984.html>.

③ 孙志建. 数字政府发展的国际新趋势：理论预判和评估引领的综合 [J]. 甘肃行政学院学报, 2011 (3): 32-43.

④ The White House. Digital Government: Building a 21ST Century Platform to Better Serve the American People [EB/OL]. (2012-05-23). [2013-11-08]. <http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government-strategy.pdf>.