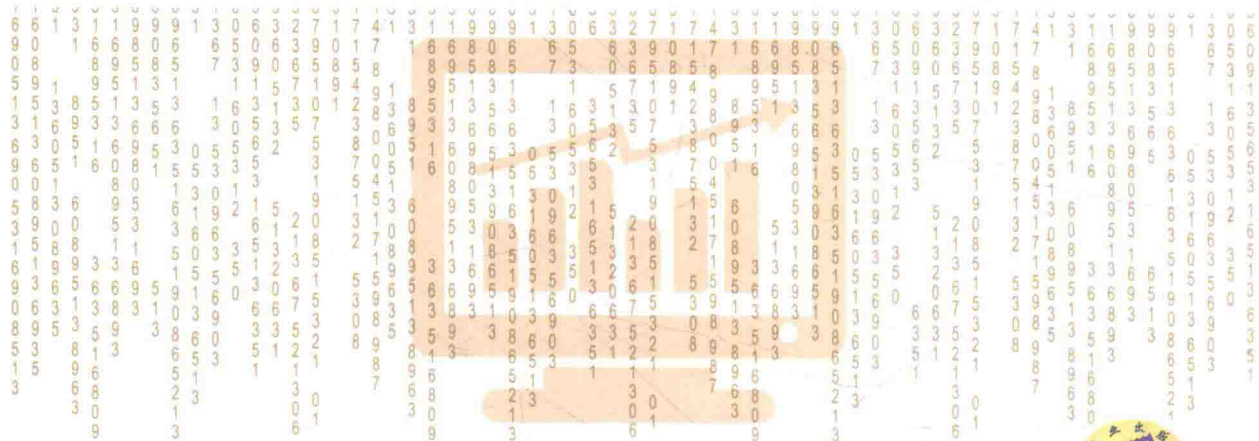




银行业信息科技风险管理高层指导委员会
银行业信息化丛书

银行数据治理

陆顾新 陈石军 王立 等编著



Data Governance of
Commercial Bank



银行业信息科技风险管理高层指导委员会
银行业信息化丛书

银行数据治理

陆顾新 陈石军 王立 等编著

Data Governance of
Commercial Bank

 机械工业出版社
CHINA MACHINE PRESS

本书通过回顾银行业信息化的历史和我国银行业数据治理的历程，阐明了数据对于银行发展的重要作用。同时，本书结合当前商业银行面临的数据治理现状，提出了银行数据治理的概念和范围，全面阐述了如何建立商业银行数据治理体系，客观提出了涵盖战略、机制、领域、技术支撑的从上至下指导、从下而上推进的多层次、多维度、多视角的全方位商业银行数据治理体系框架，即以企业战略对数据治理的目标及规划为指导，以配套的数据治理流程为推力，积极利用系统和技术手段保障数据治理的有效开展与落实。

本书可供商业银行从事数据管理的人员使用，也可供其他行业的相关从业人员参考。

图书在版编目（CIP）数据

银行数据治理/陆顾新等编著. —北京：机械工业出版社，2016.5
(2016.8重印)

(银行业信息化丛书)

ISBN 978-7-111-53485-3

I. ①银… II. ①陆… III. ①银行—数据管理—研究 IV. ①F830.49

中国版本图书馆CIP数据核字(2016)第073294号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑：王振国 责任编辑：王振国 责任校对：陈延翔

责任印制：李洋

三河市国英印务有限公司印刷

2016年8月第1版第2次印刷

184mm×260mm·13印张·321千字

3001—4900册

标准书号：ISBN 978-7-111-53485-3

定价：69.80元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务

服务咨询热线：010-88361066

读者购书热线：010-68326294

010-88379203

封面无防伪标均为盗版

网络服务

机工官网：www.cmpbook.com

机工官博：weibo.com/cmp1952

金书网：www.golden-book.com

教育服务网：www.cmpedu.com

“银行业信息化丛书”编委会

主 编：尚福林

副主编：郭利根

编 委：（按姓氏拼音排序）

陈天晴 陈文雄 方合英 甘 煜 谷 澍 侯维栋 李 丹
李 浩 李丽芳 李 翔 李振江 林晓轩 林治洪 潘卫东
庞秀生 曲家文 单继进 童 建 王 兵 王 健 王用生
谢翀达 许 文 薛鹤峰 于富海 张华宇 张依丽 朱鹤新

编 辑：（按姓氏拼音排序）

傅晓阳 龚伟华 何 禹 焦大光 金磐石 李 璠 李海宁
李建军 梁 峰 刘国建 刘秋万 刘子瑞 鲁 森 骆絮飞
吕仲涛 牛新庄 谭 波 汪 航 王 燕 吴永飞 奚力铭
徐 徽 于慧龙 余宣杰 周黎明 周天虹

工作组：（按姓氏拼音排序）

曹文中 陈宇能 黄登玺 黄绍儒 霍宝东 贾俊刚 金建新
李洪伟 李 燕 林长乐 刘文波 孙 莉 唐 宗 卫剑钊
夏建伟 闫晓鹤 张 健 张立书 钟 亮 朱学良

总 序

信息化是推动经济社会变革的重要力量。坚持走中国特色的新型工业化、信息化、城镇化、农业现代化道路，是党中央立足全局、放眼未来、与时俱进的战略决策。2014年2月27日，中央网络安全和信息化领导小组的成立，更加体现了中央保障网络安全、推动信息化发展、维护国家利益的决心。银行业作为国家经济体系的重要行业之一，是信息化的重要推动主体、参与主体和受益主体。银行业持之以恒地贯彻落实国家信息化战略，不仅是推动加快我国信息化进程的必然要求，也是银行业改革发展、转型升级和更好服务实体经济的内在需求。

近年来，我国银行业审时度势、积极作为，坚持基础建设与科技创新并重、提升服务与保障安全并举的科学发展导向，以推进信息化为契机，调整经营理念、优化经营机制、完善服务模式，在服务手段信息化、管理模式信息化、信息安全保障等方面取得积极进展，推动了银行业的核心竞争力、市场适应力和贴身服务能力的进一步提升。一是服务手段信息化发展迅速。电子银行、自助银行、智能支付终端等信息化服务渠道日渐普及，使得金融服务覆盖面更加广泛、服务方式更加便捷、服务产品更加丰富。二是管理模式信息化迈出实质性步伐。注重依托核心数据库、运用先进数据挖掘分析工具，推进银行经营决策逐步智能化，风险管理日趋精细化，产品创新逐渐体现个性化，银行业经营管理信息化水平不断提升。三是信息安全保障取得积极进展。银行业信息安全越来越受重视，相关科技基础设施建设步伐加快，多层次、立体化、全方位的信息安全保障体系正在逐步形成。

当然，我们也应该清醒地认识到，银行业信息化面临着复杂的内外部环境，核心技术受限、网络安全威胁、隐私保护和信息保密等挑战将长期存在，银行业自身认识不尽到位、技术储备不够充分、资源投入相对不足、过度依赖外包等问题仍较为突出，针对银行业特殊需求的信息化产品、工具和方法还比较单一，缺乏应对复杂需求的灵活创新能力。总的看来，银行业信息化还有很长的路要走，信息科技风险将成为当前和未来较长时期银行业的重要风险领域之一。

银行业信息化既不能因为成绩而骄傲自满，也不可因为差距而妄自菲薄，更不可因为困难而畏首畏尾。各银行业金融机构要勇于直面困难、主动迎接挑战，坚决按照国家信息化总体战略部署，切实坚持“自主可控、持续发展、科技创新”的基本方向，紧紧抓住信息化发展机遇，推动信息服务和信息安全再上新台阶。一是借助信息化推动银行业金融机构治理

能力现代化。积极引入先进的信息科技治理和管理理念，运用现代信息技术缓解治理中的信息不对称问题，推动流程银行建设，提高治理有效性。同时，理顺信息化建设的体制机制，加快信息化建设进程，为银行业转型发展提供有力保障。二是依托信息化推动金融服务智慧化。要充分利用互联网、移动计算蓬勃发展的大环境，积极应用大数据等新兴技术，创新思维模式，充分发挥金融数据和信息的价值，研发智能化、个性化、便捷化的产品和服务，灵活响应客户诉求，努力改善客户体验，尽力发掘潜在客户需求，增加产品和服务的吸引力，培育更为坚实的客户基础，形成新的业务和利润增长点。三是以自主创新增进安全可控能力。要坚持市场起决定作用的基本方针，探索形成以研发创新支持应用推广、以市场应用激发创新动力的良性正反馈机制。推动应用自主创新信息技术，建立自主创新信息技术落地银行业的配套机制，力争金融领域关键信息技术自主创新占比逐步提高，不断提升信息系统的开放性、灵活性和整体集约化水平。四是利用信息技术强化行业协作。要加强银行业信息化建设的统筹规划，促进信息化资源的集约共享，提升数据（灾备）中心布局的合理性，增强同业协同协作，共同应对外包集中度等风险。

为更好地推进落实银行业信息化战略，由银行业信息科技风险管理高层指导委员会指导推动，编著了“银行业信息化丛书”（简称“丛书”）。这套丛书致力于挖掘、研究、总结、提炼和传播国内外信息化最佳实践、宝贵经验和最新成果，内容涵盖银行业信息科技治理与管理、信息系统开发与应用创新、信息安全、基础设施与运行维护、信息科技监管等主要领域，可为银行业信息科技人才培养提供一些基础性、前瞻性、实用性的知识和信息。

展望未来，银行业信息化任务艰巨、时间紧迫。希望银行业在有关各方支持下，推动信息化工作更加积极主动、规范有效、科学前瞻，为我国银行业持续健康发展、提升服务水平提供坚实的支撑，为增强国家网络安全保障能力、提升信息化建设水平提供有力支持，为贯彻落实创新驱动发展战略、实现中华民族伟大复兴的中国梦做出积极贡献。

尚福林

序

我国银行业信息化已发展了 30 多年，银行早期的数据基本上是交易的副产品，很少用于分析决策、风险控制等。到了 21 世纪，越来越多的商业银行开始利用数据进行更为精准的客户营销、提高客户的服务水平、进行实时或准实时的风险分析和反欺诈分析，数据价值将逐步得以体现。但是，数据使用过程中暴露出来的问题也越来越突出，例如：企业层级的通用业务规则缺乏，导致不同业务部门之间的标准常有矛盾或是混淆；缺乏企业整体的宏观设计，仍然停留在面向功能为主的架构设计；数据的交换和共享困难，存在严重的数据不同步、不一致；没有明确的角色、职责、职权分工，缺乏跨部门的数据管理体系等。

事实上，一旦缺乏正确的数据管理和高质量的数据，那么数据价值就无法得到体现，进而银行业信息化建设就成了一纸空谈。因此，数据治理工作成为银行业信息化面临的重要挑战，也是银行数据价值充分发挥的前提。幸运的是，各大商业银行逐步意识到这一问题，越来越重视数据治理工作，纷纷将数据治理提上了建设议程，开始建立适应本行的数据治理体系。

数据治理是一个长期而繁琐的过程，而且面临各个方面的挑战。例如：在数据组织的一致性方面，需要建立数据标准，用于指导数据格式、数据结构、数据字典、编码规则等，这样才能保证系统间数据的有效交互；在数据质量的保证方面，不仅要技术层面建立相应的事前、事中和事后的校验、监控、矫正等机制，还要在业务层面提出合理的、一致的业务规则。当然，在技术层面建立辅助系统和工具是行之有效的手段，可以使数据变得有序、合理、规整，对银行业的风险管理、营销支持、决策分析等经营管理起到关键性作用，但最重要的还是要从战略层面上予以支持和指导，从战术层面建立具体的章程，并且作为一种重要的企业文化来执行。

在银行业信息科技风险管理高层指导委员会的牵头组织下，课题组编写了《银行数据治理》这本书，就如何建立商业银行数据治理体系进行了探讨，从战略、文化、组织、制度章程、系统工具等方面进行了详细阐述，涵盖了数据标准、元数据、数据分布、数据质量、数据安全等数据治理领域。希望这本书能对我国商业银行数据治理工作带来一定启发，从而起到重要的引导作用。

作为拥有 30 多年工作经验的银行从业人员，目睹了中国工商银行一个又一个系统的建

设过程，见证了各个系统间数据综合利用的难度，也深知数据治理在各个商业银行都是一项艰巨而具有挑战性的工作。很高兴中国工商银行能有这个机会，集中国银行、中国邮政储蓄银行等业界同人的鼎力相助编著此书，并将商业银行数据治理工作进行系统化阐述，希望这本书能够起到抛砖引玉的效果，逐步将商业银行的数据治理工作推向成熟化。

吕仲涛

前 言

全球已进入信息高速发展的时代，信息化是当今世界发展的根本趋势，是推动经济社会变革的重要力量。我国在2006年发布了《2006—2020年国家信息化发展战略》，对我国的信息化发展进行了战略指导。

为贯彻国家信息化发展战略要求，推进落实银行业信息化战略，在银行业信息科技风险管理高层指导委员会指导下，我们着手编写了“银行业信息化丛书”之《银行数据治理》。《银行数据治理》的核心是阐述如何对银行数据进行全面管理，创造价值，为银行业务发展提供全面、高效的数据支撑。

商业银行在业务开展过程中积累了海量的数据，如客户数据、业务交易数据、内部管理数据、外部数据、系统日志等。这些数据是银行最为重要的资产之一，是银行支持精细化管理、实现差异化服务、提升风险分析能力的基础。然而，海量的数据不仅意味着海量的机遇，也意味着无限的风险。目前，我国商业银行普遍面临着数据质量不高、数据支持决策能力不足等问题，导致数据远未发挥其应有的价值，并已经成为银行提高自身竞争力的巨大障碍，主要表现为：缺乏整体数据治理战略和文化、数据治理管理流程和职责划分不清、数据标准不统一、数据质量不高、缺乏配套的技术和系统支撑等。

如何开展数据治理，实现数据对业务的价值创造，是银行信息化建设中要破解的一道难题。《中国银行业信息科技“十二五”发展规划监管指导意见》指出：“商业银行要重点加强对数据治理的制度和流程建设，建立和完善数据治理制度体系，规范工作流程，理顺内部协作关系，提升数据质量和数据应用水平，提高数据价值创造能力。”数据治理需要一个完整的治理体系，并从战略、机制、领域、技术等方面予以支撑。

在信息化快速发展的今天，银行的发展要靠数据，数据要靠治理，治理需出成果，成果需要在实际应用中体现，而应用应与银行的整体发展战略契合，总结起来即为“银行靠数据、数据靠治理、治理出成果、成果靠应用、应用看前景”。

本书通过阐述数据治理的概念和范围，回顾我国银行业数据治理历程，结合分析各商业银行当前面临数据治理现状，全面阐述了商业银行数据治理体系，提出以战略、机制、领域、技术为支撑的从上至下指导、从下而上推进的实施原则，着重阐述建立多层次、多维度、多视角的全方位商业银行数据治理体系框架，包括但不限于：以企业战略对数据治理的目标及规划为指导，以配套的数据治理流程为推力，涵盖完整的数据治理领域，积极利用系

统和技术手段保障数据治理的有效开展与落实。

全书共有 13 个章节，总体内容结构如下：

第 1 章：银行的数据。阐述什么是数据以及数据对银行的作用，还介绍了银行的数据包括什么、有什么特点、管理上存在什么不足，从而使读者对银行的数据有一个大概的了解。

第 2 章：银行数据治理概述。阐述银行数据治理的背景、目标及价值，从数据治理的发展和现状出发，思考如何对银行的数据进行治理。

第 3 章：银行数据治理保障机制。介绍确保银行数据治理工作能够得到有效落实的保障措施，包括战略层面、组织机构、制度章程、支撑系统等各个方面。

第 4 章至第 12 章，具体介绍了银行数据治理各个核心领域（包括数据模型、元数据管理、数据标准、数据质量管理、数据生命周期管理、数据分布与存储、数据交换、数据安全、数据服务）目前存在的问题、治理难点、应对措施，使读者了解到每个领域的数据治理应该涉及什么样的工作、应该注意哪些细节，从而为实际的数据治理实施提供具体的指导方法。

第 13 章：数据治理展望。数据治理的成效要靠数据发挥的价值来体现。本章介绍了在大数据时代如何进一步加强数据治理、如何有效地利用数据，并以大数据在金融界的实际应用案例生动地说明了只有高质量的数据才能发挥数据的最大价值。

本书在编写过程中得到了银监会各级领导的大力指导，得到了银行业同行的有力支持，编写组所有人员也付出了辛勤汗水，在此对所有编写与指导人员表示衷心的感谢，包括中国工商银行的张艳、张颖、罗毅、李伟良、王涛、李甲长、王鹏、乔敏、杨明伟、肖睿、郭斯、巩垒、王超、袁如怡、高继荣，中国银行的张建游、王玮、曾玉、武晓利、刘晓波，中国邮政储蓄银行的汪航、芮亚南、张放、朱昌堆、陈元星，大连银行的石立军、于琦等。

尽管编写人员都投入了大量的精力，付出了艰辛的劳动，然而由于水平和能力所限，书中错误和疏漏之处在所难免，恳请大家批评指正。

编著者

目 录

总序	
序	
前言	
第1章 银行的数据	1
1.1 数据、信息和知识	1
1.2 银行的数据及作用	3
1.2.1 银行数据的来源及特征	3
1.2.2 数据对银行发展的重要作用	7
1.3 银行数据管理存在的不足	10
第2章 银行数据治理概述	15
2.1 银行数据治理的背景	16
2.2 银行数据治理的目标与价值	19
2.3 银行数据治理的历史与现状	21
2.3.1 国际社会数据治理发展历程	21
2.3.2 我国银行业数据治理前进之路	23
2.3.3 我国银行数据治理的现状	26
2.3.4 银行数据治理的思考	28
2.4 建立银行数据治理体系	30
2.4.1 银行数据治理保障机制	32
2.4.2 银行数据治理核心领域	33
第3章 银行数据治理保障机制	36
3.1 企业发展战略	36
3.1.1 数据治理文化	36
3.1.2 认知与行动	37
3.2 数据治理组织	37
3.3 数据治理制度章程	39
3.4 数据治理流程	41

3.5	数据治理支撑系统	42
3.6	数据治理成熟度模型	42
第4章	数据模型	45
4.1	数据模型体系	45
4.2	数据建模的方法、原则及意义	47
4.3	数据模型建模要求	48
4.4	数据模型设计实例	50
4.5	非结构化数据模型	60
4.6	数据模型主题	61
第5章	元数据管理	63
5.1	元数据管理概述	65
5.2	搭建元数据管理体系	67
5.3	主要元数据对象的管理	70
5.4	非结构化数据的元数据管理	74
5.5	案例	75
第6章	数据标准	76
6.1	银行数据标准的现状	76
6.2	银行数据标准的定义与分类	78
6.3	银行数据标准建设	80
6.4	非结构化数据的数据标准管理	86
6.5	案例	87
第7章	数据质量管理	90
7.1	商业银行数据质量存在的问题及影响	90
7.2	建立全面管控的数据质量管理体系	95
7.3	非结构化数据的数据质量管理	102
7.4	案例	103
第8章	数据生命周期管理	105
8.1	数据生命周期管理及意义	106
8.2	银行业数据生命周期管理的现状	107
8.3	建立数据生命周期管理体系	108
8.4	案例	118
第9章	数据分布与存储	120
9.1	商业银行系统数据存储管理	121
9.2	商业银行系统数据分布	126
9.2.1	总分行数据分布	126
9.2.2	商业银行系统主副本数据源分布	126
9.2.3	商业核心银行系统与外围系统间的数据分布	128
9.3	案例	128
第10章	数据交换	131
10.1	制定数据交换的标准	132
10.2	搭建数据交换体系和数据交换系统应用	135
10.3	全流程数据流管理	138

10.4 外部数据交换	139
第 11 章 数据安全	141
11.1 数据安全及重要性	141
11.1.1 数据安全的概念	141
11.1.2 数据安全的重要性	143
11.1.3 数据安全的发展历程	144
11.2 数据安全领域及管理	145
11.2.1 数据存储安全	145
11.2.2 数据传输安全	149
11.2.3 数据使用安全	151
11.3 案例	154
第 12 章 数据服务	158
12.1 银行在大数据时代的战略应对措施	158
12.2 构建大数据服务体系	161
12.2.1 结构化数据处理系统	162
12.2.2 非结构化数据处理系统	164
12.2.3 流数据处理系统	165
12.3 数据资产视图	167
12.4 大数据在银行业的应用	170
12.4.1 国内外银行业大数据应用案例	170
12.4.2 银行业大数据应用场景	171
第 13 章 数据治理展望	175
13.1 构建银行数据治理的长效机制	175
13.2 大数据时代的银行数据治理	176
13.3 大数据时代的银行数据利用	176
附录	180
附录 A 《中国银行业信息科技“十二五”发展规划监管指导意见之大型商业银行和股份制 商业银行篇》(摘录)	180
附录 B 《贷款统计分类及编码标准(试行)》(摘录)	190
附录 C 《银行监管统计数据质量管理良好标准(试行)》(摘录)	192
附录 D 推荐读物	195
参考文献	196

第 1 章

银行的数据

从 20 世纪 90 年代末开始，以中国工商银行、中国农业银行、中国银行、中国建设银行四大国有商业银行为代表的中国银行业陆续掀起了以数据大集中为主线的信息化建设浪潮。数据大集中缩小了我国银行与国外银行在信息技术上的差距，使得利用集中的数据进行分析 and 挖掘成为可能。

银行业信息化发展的重要基础就是各类数据和信息的整合利用，在未来一定时期内，银行的数据作用将更为凸显。我国银行业已越过数据大集中的信息整合的初期阶段，后续银行将会深入进行数据整合，基于统一的数据标准和数据模型，加强数据管控和治理，深入地对银行数据信息进行挖掘、分析、处理和快速利用，充分发挥数据仓库等系统在加强客户和市场分析能力，提升银行管理决策水平，提高银行核心竞争力等业务领域的推动作用。

特别是随着银行与互联网、合作方等相关领域的合作日趋密切，各类数据也以几何级数增长方式而迅速膨胀。当前，我国大型商业银行的数据量已经达到 TB 甚至 PB 以上级别，并且非结构化数据量也在快速膨胀。我国金融业正在步入大数据时代的初级阶段，并呈现快速发展的势头。数据爆炸性地增长，对银行进行数据分析和处理，以及源源不断的产品和业务创新等都带来了极大的挑战。未来银行将会是一个集数据分析、数据解读为一体的数据大行。因此，我们必须了解什么是数据，什么是银行的数据，银行数据到底有什么样的作用和存在哪些不足。

1.1 数据、信息和知识

数据、信息和知识都是社会生产活动中的一种基础性资源，都可以采用数字、文字、符号、图形、声音、影视等多媒体来表示，同时具有客观性、真实性、正确性、价值性、共享性、结构性等特点。三者之间既存在密切的内在联系，又有着显著的区别。

数据、信息和知识是知识工作者对客观事物感知和认识的三个连贯的阶段，主要表现在以下方面：

1. 数据的组织阶段

数据是一种将客观事物按照某种测度感知而获取的原始记录。它既可以直接来自测量仪

器的实时记录，也可以来自人的认识，但是大量的数据多是借助于数据处理系统自动地从数据源进行采集和组织的，比如银行借助计算机产生的客户数据。数据是最原始的记录，未被加工解释，不能回答特定问题。它仅仅反映客观事物的某种运动状态，除此以外没有其他意义。它与其他数据之间没有建立相互联系，是分散和孤立的。

根据结构的不同，数据可划分为结构化数据和非结构化数据两种。

(1) 结构化数据 简单来说，就是存储在关系型数据库中的数据，可以用二维逻辑表达实现。结构化的数据可以方便地通过计算机和数据库技术进行管理，信息量值出现在固定的位置，它能够忠实、详细地记录交易活动，是一种显性的表示。

(2) 非结构化数据 指信息的形式相对不固定，不方便使用数据库二维逻辑表达的数据，包括报表、账单、多媒体视频、网页、电子邮件、语音、图片、图像、办公文档等。银行的非结构数据中超过 1/2 来源于与客户之间的互动，都是以客户为中心产生的数据。非结构化数据信息量值的位置不固定，价值相对隐匿。

2. 信息的创造阶段

信息是对数据进行加工处理，使数据之间建立互相联系，形成回答某个特定问题的文本，以及被解释具有某些意义的数字、事实、图像等形式的信息。简单地说，信息是加工后的数据。

3. 知识的发现阶段

知识是知识工作者运用大脑对获取或积累的信息进行系统化的提炼、研究和分析的结果。知识能够精确地反映事物的本质。

数据、信息、知识所处的三个阶段形成一个螺旋上升式循环周期。人们运用相关的信息系统，对信息和相关的知识进行规律性、本质性和系统性的思维活动并创造出新的知识。之后，新的知识又开辟了需要进一步认识的对象领域，然后使人们补充获取新的数据和信息，进入新一轮的上升式循环周期。但三者又有一定区别：

(1) 具有各自突出的特性 数据具有真实性、客观性的特点；信息具有针对性、时效性、较少不确定性的特点；知识具有规律性、本质性、系统性的特点。

(2) 获取的途径不同 数据通过采集获得；信息通过加工获得；知识是实践经验的总结，是信息的结晶，是通过实践和学习获得的。

(3) 对知识工作者的支持过程不同 知识工作者运用数据处理系统进行数据的编码、采集、录入、存储和处理，这是数据的组织过程。知识工作者应用信息系统进行信息的捕获、导入、存储、加工、挖掘、传输、使用和支持，这是信息的创造过程。知识工作者在智能化的信息系统支持下，运用大脑进行知识的获取、提炼、分析、研究和发现，这是知识的发现过程。

在已有知识的基础上，对信息进行分析、对比、演绎并找出有价值的部分，并将其深化到已有知识架构

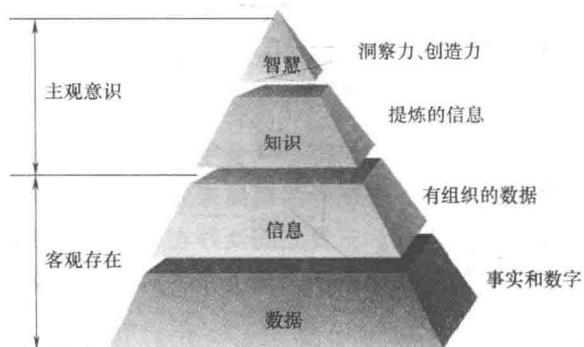


图 1-1 数据、信息、知识及智慧示意图

中，则上升为智慧。本书主要对数据以及数据产生信息进行探讨。图 1-1 为数据、信息、知识及智慧示意图。

1.2 银行的数据及作用

银行的迅速发展是建立在其庞大的数据群及基于其抽象出的信息基础之上的。如果没有数据，没有数据提炼的信息，那么银行业的发展只能成为一纸空谈。

管理大师汤姆·彼得斯说过：“一个组织如果没有认识到管理数据和信息如同管理有形资产一样极其重要，那么它在新经济时代将无法生存。”数据给商业银行创造了能够更好地理解客户、服务客户的机会。

世界经济论坛报告认为：“大数据为新财富，价值堪比石油。”大数据之父维克托曾在 2013 年预测，数据列入企业资产负债表只是时间问题。

那么，数据是什么？银行又有怎样的数据？银行的数据有什么样的特征？银行的数据到底能够发挥什么样的作用？这些都是我们关心的问题。

1.2.1 银行数据的来源及特征

我国商业银行通过广泛的渠道及丰富的产品，在业务开展过程中积累了大量的数据，如客户数据、业务处理数据、内部管理数据、外部数据、系统日志等。特别是近年来，随着社交网络、微博、微信、论坛的兴起，数据来源也已从传统的结构化数据逐渐扩展到以网络日志、社交媒体为代表的半结构化和非结构化数据，银行数据量呈现几何级数的爆发式增长，商业银行逐步进入大数据时代。

大数据时代已经悄然而至，认同这一观点的人越来越多。那么银行的数据到底从哪里来？银行有哪些数据？下面将从技术角度加以解惑，充分介绍它作为新时代变革力量的根本原因。

1. 银行的数据是什么

银行业数据原来是金融产品、业务处理、运营管理以及客户服务的副产品，只在后台起支撑作用。随着信息科学技术进步和数据处理能力的提高，大量的数据以结构化或非结构化的形式在 IT 信息系统中传输、储存、转换、分析和挖掘，并产生能够为特定业务领域服务的信息。

图 1-2 所示为银行的信息与数据。

随着网络技术的发展，特别是 Internet（因特网）和 Intranet（内部网）技术的飞速发展，商业银行在使用互联网技术、不断提升客户服务水平和加强风险管理的同时，存款、贷款及汇款（简称存贷汇）等核心业务的结构化数据与日俱增，而电子商务、移动应用、社交网络日益活跃，导致电话语音、网上银行交易记录、电子影像、网点视频、扫描文件、微博、即时通信等非结构化数据（图 1-3）、半结构化数据也日趋增大，非结构化数据的增长幅度远远大于结构化数据。

2. 银行数据的来源和分类

银行的数据主要来源于商业银行自身业务运营过程中积累的大量客户基本资料、客户交

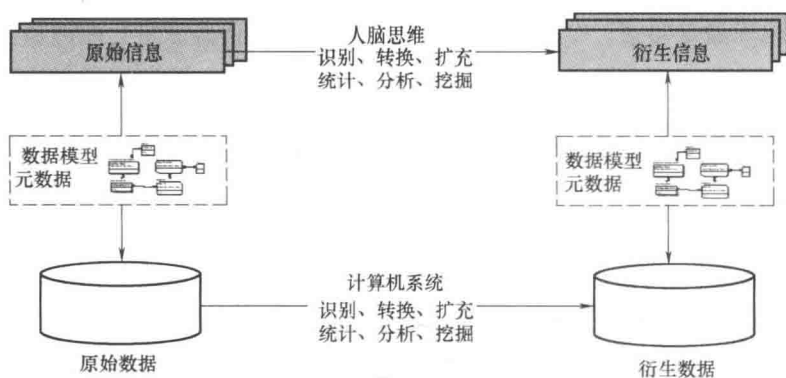


图 1-2 银行的信息与数据

易和产品服务数据、运营管理数据，当然也包括外部宏观经济环境的运行指标等数据。根据来源不同，银行信息系统的数据库可分为内部数据和外部数据两大类，见表 1-1。

(1) 内部数据

1) 客户数据：即当事人基本信息，主要用于描述客户的自身特点，属于基础数据范畴。客户（当事人）是指与银行有联络或与银行有利害关系，以及银行希望保留其信息的所有相关参与者。其中也包括银行本身，如个人、外部机构、内部机构。

① 个人客户数据包括：客户姓名、性别、国籍、证件类型、证件号码、联系地址、通信电话、职业状态、工作单位、职务、宗教信仰、婚姻状况、文化程度、子女数量、父母姓名、语言偏好，以及客户与账户的关系等信息。

② 公司客户数据包括：机构名称、机构代码、经营范围、经营状态、所属国家、所属省份、公司网站、公司地址、公司类别、法人电话、财务电话、注册资金、行业代码、经济性质、企业规模、资产规模、建筑物产权、办公面积、营业面积、客户来源等。

银行的客户数据来源于银行自身的多个系统，如客户管理系统、核心银行系统、交易系统、中间业务系统、资产负债管理系统等。银行通过客户数据，除了提供基本的金融服务外，还可以根据客户特性提供个性化的金融产品服务，同时也可以同步检验产品的占有率和推广效果等。

2) 交易及运营数据：指客户通过银行各渠道完成金融服务的相关数据，包括协议数据、介质数据、产品数据、渠道数据及事件数据。银行利用上述数据进行客户消费行为分析、金

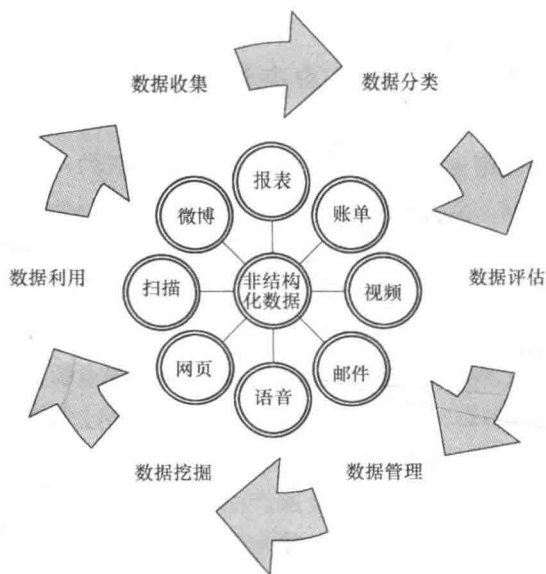


图 1-3 非结构化数据