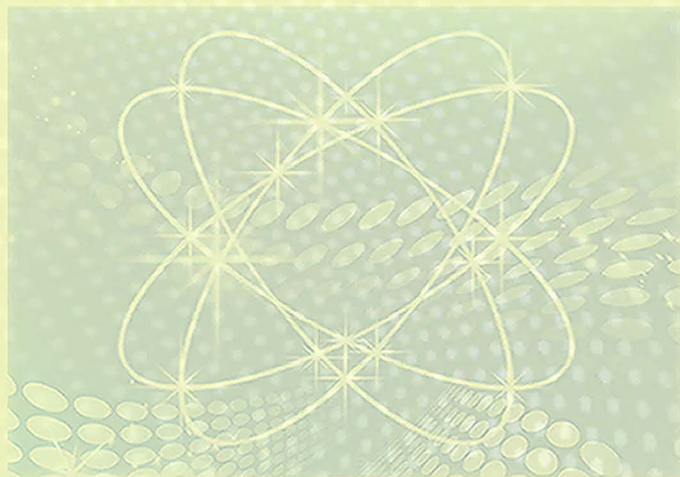


# 基于大数据环境下的档案管理基础与实务研究

常新军 吴昱达 甄杨 著



文化发展出版社

# 基于大数据环境下的 档案管理基础与实务研究

常新军 吴昱达 甄 杨◎著



文化发展出版社  
Cultural Development Press

· 北京 ·

## 图书在版编目 (CIP) 数据

基于大数据环境下的档案管理基础与实务研究 / 常新军, 吴昱达, 甄杨著. — 北京: 文化发展出版社, 2023.12

ISBN 978-7-5142-4270-6

I. ①基… II. ①常… ②吴… ③甄… III. ①档案管理—研究 IV. ①G271

中国国家版本馆 CIP 数据核字 (2024) 第 025131 号

## 基于大数据环境下的档案管理基础与实务研究

常新军 吴昱达 甄 杨 著

出版人: 宋 娜

责任编辑: 袁兆英

责任校对: 岳智勇

责任印制: 邓辉明

封面设计: 谢婉莹

出版发行: 文化发展出版社 (北京市翠微路 2 号 邮编: 100036)

网 址: [www.wenhuafazhan.com](http://www.wenhuafazhan.com)

经 销: 全国新华书店

印 刷: 天津和萱印刷有限公司

开 本: 710mm × 1000mm 1/16

字 数: 215 千字

印 张: 10.75

版 次: 2023 年 12 月第 1 版

印 次: 2023 年 12 月第 1 次印刷

定 价: 72.00 元

I S B N: 978-7-5142-4270-6

◆ 如有印装质量问题, 请电话联系: 010-58484999

# 前 言

档案管理是一种常见的信息管理方式，档案管理不仅需要对档案进行收藏和存储，也需要对档案进行分类管理，从而给用户提供更快捷方便的信息检索和档案借阅。档案资料是无时无刻不在增加的，对档案管理员来说，档案的数量是非常大的，因此其工作量也相当大，而且档案的管理是一种对资料重新整理和分类的过程，其中涉及大量重复性的细节工作，对工作的准确性和对档案的处理速度有很高的要求。如果不能满足准确性，用户不能检索出所需要的资源，那么档案管理就失去了意义；如果不能及时给用户提供所需要的资源，也就无法为用户带来良好的体验。

在信息化时代的发展背景下，以往传统落后的手工档案管理方式已经无法跟上现代档案管理相关工作的前进脚步，尤其是在档案管理技术发展日新月异的今天，现代化综合档案管理平台的广泛应用也为现代档案管理工作提供了有效的保证。另外，伴随各单位现代信息化水平的不断提高，越来越多的办公自动化系统也催生出了海量的电子档案相关资料。因此，如果不能及时、有效地对这些电子档案资料进行管理，那么就可能会发生由于文档资料丢失而无法对其完整性进行保证的情况。

本书第一章为大数据概述，主要论述了大数据的概念，大数据的发展趋势，大数据的主体分类、应用及需求三个方面的内容。第二章为档案管理工作概述，主要讲述了档案的形成与发展、档案的分类、档案工作的内容与性质、档案工作的要求与意义四个方面的内容。第三章为大数据环境下的档案信息资源整合，主要从档案信息资源整合的基本内容、档案信息资源整合的目标取向、档案信息资源整合的实施措施三个方面进行了分析。第四章为大数据环境下的档案信息资源开发利用，分别从档案信息资源开发利用的原则、档案信息资源开发利用的制约因素、档案信息资源开发利用的具体措施三个方面进行叙述。第五章为大数据环境下的档案信息服务创新性研究，主要讲述了档案信息服务研究的现状、档案信息服务研究的主要内容、档案信息服务模式面临的机遇与挑战、档案信息服务应用实例分析四个方面的内容。第六章为大数据环境下的档案信息化保障体系建设，



分别讲述了宏观管理保障体系、标准规范保障体系、信息安全保障体系、人才队伍保障体系、信息技术保障体系五个方面的内容。第七章为大数据环境下的智慧档案馆建设，主要介绍了智慧档案馆概述、我国部分城市智慧档案馆的建设、智慧档案馆的风险分析、建设智慧档案馆存在的问题及策略、智慧档案馆的发展前景五个方面的内容。

在撰写本书的过程中，作者得到了许多专家学者的帮助和指导，参考了大量的学术文献，在此表达真诚的感谢。本书内容系统全面，论述条理清晰、深入浅出，但由于作者水平有限，书中难免会有疏漏之处，希望广大同行及时指正。

本书共 21.5 万字，其中常新军撰写第一、二、三、六章共 10.5 万字，吴昱达撰写第四章和第七章一、二、三节共 5.5 万字，甄杨撰写第五章和第七章四、五节共 5.5 万字。

常新军

2023 年 5 月

# 目 录

<b>第一章 大数据概述</b> .....	1
第一节 大数据的概念 .....	1
第二节 大数据的发展趋势 .....	3
第三节 大数据的主体分类、应用及需求 .....	9
<b>第二章 档案管理工作概述</b> .....	17
第一节 档案的形成与发展 .....	17
第二节 档案的分类 .....	21
第三节 档案工作的内容与性质 .....	23
第四节 档案工作的要求与意义 .....	25
<b>第三章 大数据环境下的档案信息资源整合</b> .....	40
第一节 档案信息资源整合的基本内容 .....	40
第二节 档案信息资源整合的目标取向 .....	42
第三节 档案信息资源整合的采取措施 .....	43
<b>第四章 大数据环境下的档案信息资源开发利用</b> .....	50
第一节 档案信息资源开发利用的原则 .....	50
第二节 档案信息资源开发利用的制约因素 .....	54
第三节 档案信息资源开发利用的具体措施 .....	60



<b>第五章 大数据环境下的档案信息服务创新性研究</b> .....	69
第一节 档案信息服务研究的现状 .....	69
第二节 档案信息服务研究的主要内容 .....	71
第三节 档案信息服务模式面临的机遇与挑战 .....	80
第四节 档案信息服务应用实例分析 .....	84
<b>第六章 大数据环境下的档案信息化保障体系建设</b> .....	90
第一节 宏观管理保障体系 .....	90
第二节 标准规范保障体系 .....	95
第三节 信息安全保障体系 .....	100
第四节 人才队伍保障体系 .....	110
第五节 信息技术保障体系 .....	117
<b>第七章 大数据环境下的智慧档案馆建设</b> .....	128
第一节 智慧档案馆概述 .....	128
第二节 我国部分城市智慧档案馆的建设 .....	135
第三节 智慧档案馆的风险分析 .....	138
第四节 建设智慧档案馆存在的问题及策略 .....	142
第五节 智慧档案馆的发展前景 .....	159
<b>参考文献</b> .....	164

# 第一章 大数据概述

随着云时代的到来，大数据也吸引到了越来越多的关注。本章为大数据概述，主要论述了大数据的概念，大数据的发展趋势，大数据的主体分类、应用及需求三个方面的内容。

## 第一节 大数据的概念

大数据，也就是海量数据，一般指所含的数据集规模巨大，现在大众的软件工具无法在合理的时间进行采集、存储、分析管理的数据信息。因其在各个行业的广泛应用，使之关注热度历年来居高不下。作为人们获得新的认知、理念和创造价值的源泉，大数据的数据来源囊括我们在日常生活中可以普遍见到的网页上的图像、视频、录音，高速公路上车辆的收费记录、日常监控录像，医院的治疗病例、高端的基因测序，天文学中通过望远镜收集的信息数据等。

大数据最早出现在何时、何地？通过对大量文献资料追踪溯源，作者发现这个词出现的最早时间在 1980 年的美国，著名的未来学家托夫勒在其所著的《第三次浪潮》中将大数据热情地称颂为“第三次浪潮的华彩乐章”。从 1980 年到 2023 年，尽管大数据的发展已有 40 多年的时间，但目前对于大数据仍没有一个统一的、完整的、科学的定义。

### 一、狭义的大数据概念

受早期研究者将数据作为一种工具的思想的影响，很多研究机构和学者将大数据作为一种辅助工具，或者从其体量特征来进行定义。

作为大数据研究讨论先驱者的麦肯锡咨询公司，在 2011 年的大数据研究报告中根据大数据的数据规模来对其诠释，它给出的定义是，大数据指的是规模已

经超出了传统的数据库软件工具收集、存储、管理和分析能力的数据集。需要指出的是，麦肯锡咨询公司在其报告中强调，并不是超过某一个特定的数据容量的数据才能被定义为大数据，因为随着技术的不断进步，其数据容量也会不断地增长，行业的不同也会使大数据的定义不同。

日本野村综合研究所的高级研究员城田真琴在其专著《大数据的冲击》中通过对大数据的起源进行探讨，对大数据做出如下定义：大数据指的是运用现有的一般技术难以进行管理的大量数据的集合。

简以概之，对于大数据的狭义理解，研究者大多从微观的视角出发，将大数据理解为当前的技术环境难以处理的一种数据集；而在宏观方面，研究者目前还没有提出一种明确的看法，但多数都提出了对大数据的宏观理解。需要注意大数据在不同行业领域的差异，以及随着技术进步，其数据容量不断增长的特点。

## 二、广义的大数据概念

广义的大数据是以对大数据进行分析管理，挖掘数据背后所蕴含的巨大价值为视角，对大数据的概念进行定义。

维基百科给大数据下了这样一个定义：大数据又叫巨量数据、大资料，就是参与其中的数据量庞大得不能用现在的技术软件与工具来截取、管理、加工成需求者所需的数据。

被誉为“大数据时代的预言家”的维克托·迈尔-舍恩伯格在其专著《大数据时代：生活、工作与思维的大变革》中将大数据定义为：大数据是人们获得新的认知、创造新的价值的源泉；大数据还为改变市场、组织机构，以及政府与公民关系服务。他还认为大数据是人们在大规模数据的基础上可以做到的事情，而这些事情在小规模数据的基础上是无法完成的。

国际商业机器（IBM）公司则是从大数据的特征出发对其进行定义，它认为大数据具有“3V”特征，即规模性（Volume）、多样性（Variety）和高速性（Velocity），故大数据是指具有容量难以估计、种类难以计数且增长速度非常快的数据。

国际数据公司（IDC）则在IBM公司的结论基础上，根据自己的研究，将“3V”发展为“4V”，即数据规模巨大（Volume），数据的类型多种多样（Variety），数据的产生、处理、分析速度加快（Velocity），数据的价值难以估测（Value）。所以，IDC认为，大数据指的是具有规模海量，类型多样，数据的产生、处理、

分析速度加快，需要超出典型的数据库软件进行管理，且能够给使用者带来巨大价值的数据集。

通过对大数据的定义进行梳理可以发现，大多数研究机构和学者是从数据的规模，以及对数据的处理方式方面来对大数据进行定义的，且多是从自身的研究视角出发，因此，对大数据的定义，可谓是仁者见仁，智者见智。

本书通过参照学术领域各个研究机构和行业企业对大数据的定义，认为大数据是指信息爆炸时代出现的巨量数据或海量数据，以及因此而导致的技术和认知观念上的系列变化。它既是数据分析、管理和加工的途径，又是发现知识的逻辑，它通过把事物量化为数据来研究和分析事情。大数据以其客观性和可靠性不仅成为人们对事物认知的全新方式，也成为创新发现的全新手段。

## 第二节 大数据的发展趋势

### 一、大数据的价值及其战略意义

#### （一）大数据的价值

如今是社会高速发展、技术发达、信息畅通、人与人沟通日益紧密、生活日益便利的时代，大数据正是这一高科技时代下的产物。阿里巴巴创始人马云曾经提到过，未来的时代不会是 IT（信息技术）时代，而将是 DT（数据科技）时代，由此可见大数据多么重要。

有人形象地说，数据就是蕴藏着能源的煤矿。煤炭按其性质分为烟煤、无烟煤和褐煤等，而露天煤矿和深山煤矿的挖掘费用又不同。类似地，大数据也不在于“大”，而是“有用”。价值含量，即挖掘成本重于量。对许多行业来说，如何使用这些大规模数据，是在竞争中取胜的关键。

大数据的价值主要表现在三个方面：第一，向众多消费者提供产品或服务的企业能够运用大数据开展精准营销；第二，做到小而美的模式，中小微企业能借助大数据进行服务转型；第三，在互联网压力下，传统企业不得不进行转型，这就要求企业紧跟时代步伐，充分挖掘大数据的价值。

然而，大数据对经济发展的巨大影响并不代表它可以替代对社会问题的所有理性思考，科学发展逻辑也不应该湮没于大量数据之中。著名经济学家路德维

希·冯·米塞斯曾提醒过：“就今日言，有很多人忙碌于资料之无益累积，以致对问题之说明与解决，丧失了其对特殊的经济意义的了解。”<sup>①</sup>这确实是需要警惕的。

在这飞速发展的智能硬件时代里，困扰着应用开发者的重要难题是怎样在功率、覆盖范围、传输速率及成本等因素中寻找细微的平衡。企业组织分析及运用有关资料可降低成本，提高效益，开发新品，做出更加英明的业务决策，等等。通过大数据与高性能相结合的方式进行分析，以下这些有益于企业的事情就有可能出现：第一，及时分析故障、问题及缺陷产生的根本原因，每年可为企业节约数十亿美元；第二，规划数千辆快递车辆的实时交通路线，以免造成交通拥挤；第三，对全部SKU（存货单位）进行分析，并将利润最大化作为定价及清理库存的目标；第四，按照顾客的购买习惯向顾客推送其可能关心的优惠信息；第五，在众多的顾客中迅速确定金牌顾客；第六，采用点击流数据分析与数据挖掘相结合的方法规避欺诈行为。

## （二）大数据的战略意义

大数据技术所具有的战略意义并不是拥有海量的数据信息，而是将这些有意义的信息专业化。也就是说，若将大数据类比为—个行业，则该行业要想获得利润，关键是要提升数据“加工能力”，并通过“加工”来使数据“增值”。

在云时代到来的背景下，大数据已经引起越来越多人的注意。著云台的分析师团队表示，大数据常用来描述企业制造出的海量非结构化、半结构化的数据，当它们被下载至关系数据库中进行分析时，会耗费太多的时间与资金。大数据分析常常与云计算相联系，因为对大型数据集的实时分析要求 MapReduce（在大规模数据集上并行操作的编程模型）等框架将工作分配给几十台、几百台乃至几千台计算机。

## 二、大数据与传统数据库

大数据分析相比于传统的数据仓库应用，具有数据量大、查询分析复杂等特点。传统数据库是 GB/TB 级（计算机储存容量单位，1PB=1024TB、1TB=1024GB）的，高质量、较干净、强结构化、自上而下、重交易、确定解的数据仓库。大数据是 PB 级以上的，有噪声、有冗余、非结构化、自下而上、重交互、满意解的数据仓库。大数据出现后，大数据对数据库的高并发读写要求、

<sup>①</sup> 董良，何为凯，赵儒林．云计算技术与实现[M]．东营：中国石油大学出版社，2021．

对海量数据的高效存储和访问需求、对数据库高可扩展性和高可用性的需求，使传统 SQL（关系型数据库）逐渐没有用武之地，NoSQL（非关系型数据库）这一模式变得非常流行。互联网巨头对 NoSQL 数据模式的应用非常广泛，如谷歌的 Big Table（分布式数据存储系统）、脸书的 Cassandra（开源分布式 NoSQL 系统）、甲骨文公司的 NoSQL 等。虽然大数据的存储和处理已有了成熟的解决方案，但是在系统软件中占较大比重的操作系统却没有太大变化，一些重要的技术问题还没有解决，如操作系统对新兴计算资源的直接抽象的调度、分布式文件系统下的统一数据视图、全数据中心范围内的能耗管理、大数据下的安全性等相关技术还不成熟，需要进一步研发。

### 三、大数据与互联网

大多数研究大数据的商业公司都有明确的商业目的，即更好地支撑万维网（Web）服务，如谷歌搜索引擎服务、社会化网络服务（SNS）网站、新浪微博网站等。在大数据的驱动下，Web 服务可以提供更加流畅的网页交互体验和更加快速的社会资讯获取，让人们的日常工作和生活更加便捷，实现更加深入的人、机、物融合。

回顾一下 Web 的发展，Web1.0 时代的 Web 内容主要由网站服务商提供。Web2.0 时代，用户大量参与 Web 内容的生产，如博客和微博内容。Web3.0 时代，人、机、物将共同参与，使 Web 形成对真实世界的全面映射。

### 四、大数据对社会发展的影响

大数据用户行为反映其真实需求。一切行为皆有前兆，未来的不确定性是人类产生恐惧的根源之一，而大数据技术在一定程度上可以帮助我们预测未来。简单来说，大数据有帮助人们在各种数据中迅速获得宝贵信息的能力。在大数据时代，软件的价值表现为它所带来的数据的大小、流量和活性；企业的价值体现在它所掌握资料的大小、活性及其搜集、利用资料的能力等方面，而这一切都决定着企业核心竞争力的强弱。就国家层面而言，国家数字主权表现为对数据的拥有与支配。数字主权必将成为边防、海防和空防后的又一个大国博弈的空间。

#### （一）泛互联网化

泛互联网化是收集用户数据的一种低成本方式，能够为人们提供数据规模和数据活性。泛互联网化给软件的使用带来三个变化：跨平台化、门户化和碎片化。



### 1. 跨平台化

应用软件深度整合网络浏览器功能，使移动终端（手机、平板电脑）拥有相同的体验和协同的功能。

### 2. 门户化

用户无须启用其他软件即可完成绝大多数的工作并满足沟通需求，对于个性化的用户需求，可以直接调用第三方应用或插件来满足需求。

### 3. 碎片化

把原来大型臃肿的软件拆分成多个独立的功能组件，用户可以按需下载使用。这三个特征的核心意义分别在于收集用户行为数据，提高客户黏性及降低软件总体拥有成本、改变商业模式。

## （二）行业垂直整合，越靠近消费端的企业越有发言权

开源软件加速了基础软件的同质化趋势，而软、硬件一体化的趋势进一步弱化了产业链上游的发言权。大数据产业结构发展趋势有两个维度：第一个维度是大数据产业链，围绕数据的采集、整理、分析和反馈；第二个维度是垂直的行业，如媒体、零售、金融服务、医疗和电信等行业。

从这两个维度来看，大数据有三类商业模式：大数据价值链环节，专注于价值链的高附加值；垂直产业的大数据整合，利用大数据提高垂直产业效率；大数据使用者，提供大数据基础设置、技术和工具。

## （三）数据成为资产

未来企业的竞争将是拥有数据规模和活性的竞争，以及对数据解释和运用的竞争。大数据已经成了企业与社会共同瞩目的重要战略资源之一，也成为人们竞相抢占的一个新热点。因而，企业必须提前制订营销战略计划，抢占先机。以数据为中心，可演绎为租售数据模式、租售信息模式、数据媒体模式、数据使用模式、数据空间运营模式及大数据技术提供商六种新型商业模式。下面简要介绍前四种模式。

### 1. 租售数据模式

简单来说，租售数据模式就是租 / 售广泛收集的、精心过滤的、时效性强的数据。

### 2. 租售信息模式

租售信息模式通常着眼于某一产业，大量搜集相关资料，对信息进行深度提

取和整合，是利用海量数据中心加专用传播渠道来服务某一产业的模式。这里的信息是指经加工处理并携带某种行业特征数据的集合。

### 3. 数据媒体模式

把行业定位在媒体，运用数据挖掘技术来帮助顾客打开精准营销的大门，而企业的收益则来源于顾客增值部分的提成，这种模式就是数据媒体模式。该模式通常用来帮助银行和运营商等数据大户发展新业务。

### 4. 数据使用模式

数据使用模式是把产业定位于特定行业，经过海量的数据支持，挖掘和分析数据之后预测出有关主体行为从而进行业务的一种模式。数据使用模式是根据大数据技术进行高收益、低风险经营，给企业带来了全新盈利的模式。今后会产生更多让能型数据让能业务模式，这些让能模式具有创新业务特质。该模式的主要成功要素在于保持资料的真实性与完整性，以及在适当时候对风险的分析。数据愈完善，风险愈低，对确保企业高收益就愈有利。

## （四）数据管理成为核心竞争力

数据管理成了企业的核心竞争力并对企业的财务表现产生了直接的影响。随着“将数据资产作为企业的核心资产”理念深入人心，企业对数据管理有了更加明确的理解，把数据管理作为核心竞争力并不断发展，战略性地规划和利用数据资产就成了企业数据管理的中心内容。数据资产管理效率对主营业务收入增长率和销售收入增长率均有显著的正向影响；另外，对互联网思维下的公司来说，数据资产竞争力占据了相当一部分的份额，数据资产管理效果会对公司财务表现产生直接的影响。

## （五）数据泄露事件频发

今后数年内，数据泄露事件可能以成倍的速度增长，除非这些数据能从源头上获得安全保障。可以这样认为，未来财富 500 强的公司都将面对数据攻击，而不管其是否进行了安全防范。如今的企业，无论规模大小，都需要重新审视安全的定义。财富 500 强企业大都将设立首席信息安全官。企业要有新举措才能保证自己和顾客的数据安全，一切资料从一开始就必须得到安全保障，安全防范并不是资料保存过程中的最后一环，单纯强化最后一项安全措施是无济于事的。

## （六）数据生态系统复合化程度加强

大数据世界不仅仅是单一而庞大的计算机网络，它是由众多活动构件和多元

参与者要素组成的生态系统，是由终端设备提供商、基础设施提供商、网络服务提供商、网络接入服务提供商、数据服务使能者、数据服务提供商、联系人服务、数据服务零售商和其他许多参与者组成的一个生态系统。如今数据生态系统已初具雏形，下一个阶段将进行系统内部角色的细分（市场的细分）、系统机制的调整（商业模式的创新）、系统结构的调整（竞争环境的调整）等，使得数据生态系统复合化的程度逐步加深。

## 五、云平台数据更加完善

企业越来越希望能将自己的各类应用程序及基础设施转移到云平台上。就像其他信息技术系统那样，大数据的分析工具和数据库也将走向云计算。

云计算可以给大数据带来怎样的改变？

首先，云计算给大数据带来了可弹性扩展且价格比较低廉的存储空间与计算资源，让中小企业同样能够像亚马逊那样借助云计算完成大数据分析。2013年以来，大数据技术已经开始与云计算技术进行紧密的结合，二者的联系有望在今后变得更加紧密。在此基础上，物联网和移动互联网也会齐头并进，共同推动大数据革命的发展，使大数据营销产生更加重要的作用。

传统 IDC 在历史上一直是这样一种格局，互联网巨头也纷纷提供这类业务。但是最近网盘发展势头迅猛，在大数据视角下，各个商家都捕捉到了大数据的商业机会，并开始抢夺个人、公司等数据资源。海外 Dropbox（网络文件同步工具）、国内微盘等就是这类企业的典型代表。此类企业的发展追求是能够发展成为数据聚合平台并转变为多元化盈利模式。

从数据量上看，非结构化数据是结构化数据的 5 倍及以上，任何一个种类的非结构化数据处理，都可以重现现有结构化数据的辉煌。大规模、高速发展的企业可能会在语音数据处理、视频数据处理、语义识别、图像数据处理等方面崭露头角。

其次，云计算 IT 资源规模大，分布更广，为异构系统众多的公司提供了一种对数据进行及时和精确处理的强大手段，甚至可以说该手段是唯一手段。

当然大数据迈向云计算也依赖于数据通信带宽与云资源池的提升，保证原始数据能够迁移至云环境中，资源能够随着需求弹性拓展。

### 第三节 大数据的主体分类、应用及需求

#### 一、大数据的主体分类

在日益发展的信息时代，人们的生活、工作已经离不开网络了，上网记录也成了数据。互联网中的大数据不易归类，百度将数据划分为用户搜索生成的需求数据和从公共网络中获得的信息；阿里巴巴将这些信息按照商业价值划分为交易、社交、信用、移动等。互联网大数据既可划分为互联网金融数据，也可划分为用户的消费、用户的地理位置、用户的行为和用户的社交。

从社会宏观的角度按使用主体的不同可以将互联网大数据划分为如下三种类型。

##### （一）政府的大数据

各级政府各机关掌握着大量原始数据，其中，有各种环保、气象、电力等生活数据，道路交通、自来水、房屋等公共数据，以及教育、医疗、信贷、金融等服务数据。若将这些数据联系在一起，让它们流动起来并对其进行全面的有效管理，那么这些数据就会创造出巨大的社会价值与经济效益。

现代城市凭借网络智能在向智慧化迈进，不论是智能电网还是智慧医疗、智能交通、智慧环保，都需要大数据来支持，大数据是智慧城市最核心的本钱。智慧城市的构建，大数据能够从各个方面为决策和智力提供支持。作为国家管理者，政府应逐步公开数据，供更多胜任工作的机构组织和个人进行分析和使用，为人类造福。

##### （二）企业的大数据

企业要想快速成长，要想盈利，要想维护顾客，要想传递价值，要想支持规模，要想提高影响力，要想为买家服务，要想提升质量，要想节约成本，要想战胜对手，要想打开市场，就必须要有数据作为支撑。企业要想为迅速增长的消费者群体提供差异化产品或服务并进行精准营销，就必须借助于大数据。网络企业要依托大数据进行服务升级和方向转型，传统企业也要在无所不在的压力下寻求改变，整合新技术并不断向前发展。

在信息技术不断发展的今天，数据已经成为企业最核心的资产与最基本的要

素，它已经成为一个行业，并壮大成了一个供应链模式，慢慢地链接成了一条贯穿始终的数据供应链。在互联网时代，外部数据之间相互自由连接的重要性已经逐步超越了内部数据的单一性，而单个企业内部数据则更难与互联网整体数据相比拟。综合所提供的数据，推动数据应用、推动数据整合加工的新型公司具有明显的竞争优势。

在大数据时代催生了极具影响力的互联网企业的同时，传统 IT 公司纷纷进军互联网领域，通过移动云计算和大数据技术对产品进行改进、对平台进行升级、对企业进行改造，两类企业之间相互学习、合作、竞争。

### （三）个人的大数据

每个人都可以在互联网上建立自己的信息中心来对所有个人大数据信息进行累计、记录、收集和存储。按照有关法律，经本人亲自许可，一切与个人有关的信息都会变成宝贵的资料，由第三方收集后可迅速加工并得到个性化数据服务。运用信息技术的各类可穿戴设备（包括置入的多个芯片）均可通过感知技术获取个人大数据，其中包括但不限于体温和心率，视力和其他类型的身体数据，社会关系、地理位置和购物活动的社会数据。个人可选择向医疗服务机构授权让他们查阅身体数据以监控目前身体状况并制订私人健康计划；也可以授权专业金融理财机构让其看个人金融数据来做相关理财规划和收益预测。当然，国家相关部门也会对其实施法律许可的预防监控来维护公共安全和防止犯罪。

个人大数据严格受法律保护，其他第三方机构须依法经本人许可后方可使用，并对数据须进行公开、透明、综合的监督；收集个人数据应清楚地根据国家法律规定，收集内容和范围应由使用者自行决定；数据只能由用户明确授权才能严格处理。

## 二、大数据的应用

### （一）大数据在个人生活上的应用

#### 1. 在各类文化产品上的应用

如何利用大数据来提升数据价值、理解消费者需求、让生产出来的文化商品能够符合目标人群需求、增加利润空间已经成为无数文化企业研究的中心问题。

大数据在影视业的运用直接迎合了受众的需求。以往多是依靠导演、主演、制片人等的经验来拍摄影视剧。在许多情况下，投资要寻求社会关系的保障；播