

但是,当一个组织、一个城市或一个国家的成员都将个人工作和生活的事物存储到计算机系统时,数据将反映这个组织、这个城市或这个国家整体的状况,包括国民经济与社会发展的各种规律和问题,这些事情是事先不知道的,即信息化工作将社会经济规律这些未知的东西也存储到计算机中。

在新型的数字产品方面,数据更是未知的。例如,电子游戏创造了一个全新的世界,这个世界的所有场景角色都是虚拟的,甚至有虚拟的货币。这些虚拟世界的事物又通过游戏玩家与现实世界联系在一起。因此,游戏世界表现出的和其内在的东西在现实世界中是没有的,是未知的。

三是多样性和复杂性。伴随着数据爆炸,越来越多的数据被存储到计算机系统中,数据的类别和数据的形式有很多种,因此计算机系统中的数据是多样的和复杂的。

数据的多样性是指数据有各种类别,如语言的、行业的、空间的、海洋的、DNA 等,也有在互联网中或不在互联网中的、公开或非公开的、企业的或政府的。数据的复杂性是指数据具有各种各样的格式,包括专用格式和通用格式,并且数据之间存在着复杂的关联性。

### (1) 数据类别

数据主要有以下类别:

① 私人数据库。这是指存储在个人计算机系统数据库,包括个人隐私数据和個人工作数据。个人工作数据内容涉及繁多,可以是工作单位的数据、个人因工作需要收集的数据和

因其他需要获得的数据等。而另一类个人数据常被忽视,即散落在互联网的个人隐私数据。

② 企业数据库。包括企业生产经营数据、客户数据、竞争对手数据、行业数据等,这些数据主要存储在企业的计算机系统中。

③ 政府数据库。这是指存储在政府计算机系统数据库。

④ 公共数据库。这主要是指存储在公共网站上的数据。这些数据能够通过搜索引擎访问。

## (2) 数据的组织形式

数据的组织形式主要有:

① 专用格式数据。有相当多的数据由专用数字化设备产生,如医学影像数据(X光、B超、CT等),还有GIS、多媒体等数据。这些数据的处理需要专门的设备或专门的软件。

② 通用格式数据。在信息化早期,大多数数据被存储在通用数据库中,由通用的数据库管理系统(如Oracle、DB2等)来管理。这些数据库结构清晰,处理方便。

③ 互联网数据。互联网上的数据,其门类和格式繁多,还包括很多数据垃圾、病毒。由于互联网数据的形成,计算机系统中的数据更加显现出自然界的一些特征。

## 3. 数据的体量

从数据的体量来看,可以归纳出四种属性。

一是规模性。这主要体现在两个维度:样本容量大,样本容

量大大超过解释变量的数目,这称为“高大数据”;变量数量多,解释变量的数量超过样本容量的数量,这称为“庞大数据”,是一种高维或超高维数据。这两种数据既给计量经济建模提供了很大的灵活性,也带来了“维度灾难”(Curse of Dimensionality)的挑战。

二是多样性。这是指大数据既有结构化数据,又有非结构化数据。非结构化数据提供了传统数据所没有的丰富信息,极大地拓展了经济学研究的边界与范围。

三是高速性。这是指高频数据甚至实时数据的可获得性。

四是准确性。这是指大数据容量很大,噪声可能很大,因此信息密度较低,这使得统计学的一些基本原理如充分性原理和降维原则,在总结、提取数据信息时就显得非常有用。同时,由于大数据结构复杂、形式多样,信号噪声通常比较大,传统的统计充分性原理和降维原则需要有所创新与发展。

#### 4. 新型生产要素

数据已被作为一种新型生产要素,从这个角度来看,可以归纳出四种属性。

一是非排他性。非排他性即非独占性,即可复制、可共享、可交换、可多方同时使用,共享增值。

二是非竞争性。非竞争性即开发成本高,在动态使用中发挥价值,边际成本递减。

三是非稀缺性。非稀缺性即万物数据化,快速海量积累,总

量趋近无限,具有自我繁衍性。

四是非耗竭性。非耗竭性即可重复使用、可组合、可再生,在合理运维情况下可永远使用。

### 三、数据和信息的关系

数据要素应用于经济和社会领域,需要经历多次处理与挖掘:原始数据的价值无法直接展现,需要将其转化为有价值的信息。因此,我们需要对数据与信息之间的关系进行分析,从基础原理的角度对数据要素如何推进经济发展进行阐释。本部分将从四个角度阐释数据和信息的关系。

#### (一) 信息学角度

##### 1. DIKW 模型

在信息学领域,DIKW 模型是经典的关于数据、信息、知识及智慧的价值链条。它最早由英国诗人托马斯·艾略特(Thomas S. Eliot)提出。1982年,美国教育学家哈伦·克利夫兰(Harlan Cleveland)在其发表的文章《信息即资源》(“Information as a Resource”)中引用了艾略特的诗句。后来,米兰·泽莱尼(Milan Zeleny)和罗素·阿科夫对这一模型不断加以扩展。

简单来说,DIKW 模型将数据、信息、知识和智慧纳入一个金字塔形的层次体系,展示上述四者在组织形式上的关系,如图 1-2 所示。