Theory and Technology of Data Engineering

# 数据工程理论与技术

戴剑伟 吴照林 朱明东 龚建华 等编著

国防工业出版社
National Defense Industry Press

# 数据工程理论与技术

戴剑伟　吴照林　朱明东　龚建华
曾昭文　许庆华　张胜　姚远　文峰　编著

国防工业出版社
·北京·

# 内容简介

本书以数据的生命周期为主线，重点研究数据建模、数据标准化、数据管理、数据应用和数据库安全有关理论和技术。

数据建模主要介绍了数据建模的理论、方法和工具；数据标准化重点研究了数据标准化的内容和数据标准化方法；数据管理介绍了数据存储、备份与容灾基础知识和基本技术，以及数据质量管理的方法；数据应用研究了数据集成、数据挖掘、数据服务、数据可视化和信息检索的方法和技术；数据库安全重点研究了数据库安全威胁和安全机制。

本书着重理论、技术和实践相结合，内容实用、覆盖面广，可作为相关专业研究生和高年级本科生的教材，也可作为工程技术人员的参考书。

**图书在版编目(CIP)数据**

数据工程理论与技术/戴剑伟等编著. —北京：国防工业出版社，2010.7

ISBN 978-7-118-06979-2

Ⅰ.①数… Ⅱ.①戴… Ⅲ.①数据管理 Ⅳ.①TP311.13

中国版本图书馆 CIP 数据核字(2010)第 126959 号

※

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

北京嘉恒彩色印刷有限责任公司

新华书店经售

*

开本 787×1092 1/16 印张 17¾ 字数 410 千字

2010 年 7 月第 1 版第 1 次印刷 印数 1—3000 册 定价 38.00 元

(本书如有印装错误，我社负责调换)

国防书店：(010)68428422 发行邮购：(010)68414474

发行传真：(010)68411535 发行业务：(010)68472764

# 前　言

数据是人类社会活动、科技活动、经济活动和军事活动的产物，它凝聚着重要的科学价值、经济价值、社会价值和军事价值，是当代社会进步、经济发展、科技创新、新军事变革的重要资源和基础，数据资源的开发和利用已成为推动社会发展和进步的重要力量。

随着信息技术的迅猛发展，我们面临许多新的数据问题。首先，数据量爆炸式增长，数据管理的难度和压力日益增加；其次，社会信息化发展导致不同的信息系统的交流日益广泛，我们对数据共享共用的要求越来越高，以业务为中心的传统信息系统开发模式难以适应这种要求；最后，很多信息系统本身功能强大，但由于缺乏有效数据的支撑，导致其应有的效益难以发挥，数据资源的建设发展已成为制约信息系统效益发挥的瓶颈。数据工程正是在这种背景下产生的一门新兴学科。

数据工程是信息系统的基础工程，围绕数据的生命周期，规范数据从产生到应用的全过程，目标是为信息系统的运行提供可靠的数据保障和服务，为信息系统之间的数据共享提供安全、高效的支撑环境，为信息系统实现互连、互通、互操作提供有力的数据支撑。它是实现这些目标的一系列技术、方法和工程建设活动的总称。主要研究内容包括数据建模、数据标准化、数据管理、数据应用和数据安全有关理论和技术。

数据建模是对现实世界中具体的人、物、活动、概念进行抽象、表示和处理，变成计算机可处理的数据，也就是把现实世界中的数据从现实世界抽象到信息世界和计算机世界。数据建模主要研究如何运用关系数据库设计理论，利用数据建模工具，建立既能正确反映客观世界，又便于计算机处理的数据模型。

数据标准化主要为复杂的信息表达、分类和定位建立相应的原则和规范，并在信息化建设中予以宣传、贯彻和执行的过程。数据标准化重点研究数据标准化的内容、数据标准化的方法等。

数据管理是保证数据有效性的前提。首先要通过合理、安全、有效的方式将数据保存到数据存储介质上，实现数据的长期保存；然后，需要对数据进行维护管理，提高数据的质量。数据管理研究的主要内容包括数据存储、备份与容灾的技术和方法，以及数据质量管理方法。

数据应用通过数据集成、数据挖掘、数据服务、数据可视化、信息检索等手段，将数据转为信息或知识，辅助人们进行决策。数据应用研究的主要内容包括数据集成、数据挖掘、数据服务、数据可视化和信息检索的相关技术和方法。

数据安全是采取一定的安全措施，确保合法用户采用正确的方式、在正确的时间对相应的数据进行正确的操作，确保数据的机密性、完整性、可用性和合法使用。数据安全包括数据访问安全、数据传输安全、数据存储安全和数据库安全。

全书共分9章。第1章为绪论，介绍数据及数据工程基本概念，分析了数据工程建设的现状与发展。第2章为数据建模，介绍关系数据库设计理论、数据模型、数据建模方法和工具。第3章为数据标准化，研究元数据、数据元、数据模式和数据分类与编码标准的内容及其标准化方法。第4章为数据存储、备份与容灾，介绍数据存储、备份与容灾基础知识和基本技术。第5章为数据质量管理，研究数据质量的描述元素及评价方法、数据质量控制过程和单源数据的清理方法。第6章为数据集成，研究数据集成的常用方法、标准和技术。第7章为数据挖掘，主要介绍数据挖掘的基本概念、方法和多维数据分析方法。第8章为数据应用，主要研究数据服务、数据可视化和信息检索的主要技术。第9章为数据库安全，重点研究了数据库安全威胁和安全机制。

在本书的编写过程中，我们查阅了大量资料，并参考和引用了许多国内外相关书刊和文献，在此衷心感谢所有参考文献的作者和因疏忽而未在参考文献中列出的作者；还借鉴和吸收了“国家科学数据共享工程”、“中国科学院科学数据库”、“国家卫生信息标准化”等项目的研究成果，对这些项目的研究人员表示诚挚的谢意。

由于作者水平有限，加之信息技术发展日新月异，数据工程理论与技术不断发展和完善，书中难免有错误与不妥之处，敬请读者批评指正。

编著者

2010年4月于武汉

# 目　录

# 第1章 绪 论

在信息化社会，对数据资源的开发和利用成为推动社会发展和进步的重要力量。随着信息化建设的不断推进，数据总量呈指数式增长，数据维护管理难度增大，但数据集成与共享的需求是越来越迫切，数据资源建设已成为制约信息系统效能发挥的瓶颈。本章主要介绍数据的基本概念，阐述数据工程产生的背景、内涵及研究对象，介绍数据工程的现状与发展。

## 1.1 数 据

### 1.1.1 数据的定义与生命周期

**1. 数据的定义**

数据(data)是对客观事物的性质、状态以及相互关系等进行记载的物理符号或物理符号的组合。比如描述5个人，可以用5、五、伍、正、101、five、☆或者条形码表示。符号可以是数字，也可以是文字、图形、图像、声音等，因此数据的类型有数值型和非数值型。数值型数据可以直接进行科学计算，使得客观世界严谨有序。非数值型数据是除了数值数据以外的其他数据，使得客观世界丰富多彩。

**2. 数据的生命周期**

数据的生命周期可以划分为数据描述、数据获取、数据管理、数据应用四个阶段，每个阶段又包括多个具体的数据活动。

(1) 数据描述。数据描述阶段是数据生命周期的开始阶段，需要对应用领域进行深入研究分析，制定出相应的数据标准，或基于成熟的数据标准，完成数据的定义，最后，通过具体的分析过程完成数据结构设计。

(2) 数据获取。数据获取阶段是数据的实际积累和完善的过程。数据获取阶段的活动包括原始数据获取、数据预处理、数据规范化处理等具体活动。一般情况下，通过原始数据获取活动得到的第一手数据，再通过数据预处理活动对数据进行预处理，去除其中非本质的、冗余的特征，最后通过数据规范化处理后，得到有效数据。

(3) 数据管理。数据管理阶段的活动包括存储管理、数据安全、数据维护、数据质量保证等具体活动。数据管理是数据有效性的重要保证，为后阶段的数据应用打好基础。

(4) 数据应用。数据应用阶段的活动是数据深加工过程，也就是数据价值的具体实现。数据应用阶段的活动可按照具体的技术特征细分为数据挖掘、信息检索、数据集成、数据可视化等活动，这些活动实际上就是数据应用过程中所使用的不同技术手段。在数据工程领域中，主要研究如何将这些具体活动应用到数据深加工过程中。

### 1.1.2 数据的特性

数据的特性，是指数据区别于其他事物的本质属性。数据的基本特性主要有客观性、共享性、不对称性、可传递性和资源性。

（1）客观性。数据是描述物质的存在、相互关系、运动状态和变化规律的，它是对客观自然现象和规律的基本理解，反映事物的本质，是客观存在的。

（2）共享性。数据区别于物质、能源的一个重要特征是它可以被共同占有、共同享用。根据物能转化定理和物与物交换原则，得到一物或一种形式的能源，会失去另一物或另一种形式的能源。而数据交换双方不仅不会失去原有的数据，而且还会增加新的数据。

（3）不对称性。数据的不对称性可从两个方面理解：首先是对客观事物的认识，不同人（或者说对事物认识的主体）有不同的认识程度，因而对某一个客体所获取的数据不尽相同，就造成了对这个客观事物产生了不同的认识或者说不完全相同的认识；另一种情况是反映客观事物的数据，不能被不同人完全一致地占有，某些人占有的多，某些人占有的少，这就造成了同一事物的数据在不同群体（或人）中的差异，造成了不对称。由此会产生人们对同一事物的不同认识，当然也就会产生不同的结论。

（4）可传递性。数据依靠各种传播工具实现传递，它可以在不同载体之间、不同区域之间进行传递，在传递过程中数据可能一成不变，也可能产生了数量的增减或价值的变化。数据在传递过程中不断表现出它的价值。

（5）资源性。人类进入了21世纪，信息成为继物质和能量之后的第三大资源，而信息的产生是以数据为基础的，所以数据的资源性特征是显而易见的。

### 1.1.3 数据与信息、知识、智慧的关系

**1. 数据与信息的关系**

将数据放到一个语境（context）中，给予它一定的含义，就成为信息，简单地说，信息 = 数据 + 语境。信息普遍存在于自然界、社会以及人的思维之中，是客观事物本质特征千差万别的反映，信息是对数据的有效解释，信息的载体就是数据。数据是信息的原材料，数据与信息是原料与结果的关系。

例如，“6000”是未经加工的客观事实，它是数据，如果将“6000”放到特定的语义环境中，如“6000米是飞机的飞行高度”，它就是信息，再比如“8000”是数据，而“8000米是山的高度”就是信息。

**2. 信息与知识的关系**

知识是人们对客观事物运动规律的认识，是经过人脑加工处理过的系统化了的信息，是人类经验和智慧的总结，简单地说，知识 = 信息 + 判断。信息是知识的原材料，信息与知识是原料与结果的关系。

例如，人们将飞机飞行高度与山的高度两条信息之间建立一种联系，加上自己的判断就产生了知识，比如“如果飞机以6000米的飞行高度向高度为8000米的高山飞去，飞机就会撞毁”就是知识。

**3. 知识与智慧的关系**

在了解多方面的知识之后，能够预见一些事情的发生并采取行动，就是智慧，简单地

说，智慧 = 知识 + 整合。知识是智慧的原材料，知识与智慧是原料与结果的关系。人类的智慧反映了对知识进行组合、创造及理解知识要义的能力。

例如，根据“如果飞机以6000米的飞行高度向8000米的高山飞去，飞机就会撞毁”这条知识，可以预见飞机撞山的发生，并采取行动，“让飞机始终保持在高于山的高度飞行”，这就是智慧。

综上所述，数据、信息、知识、智慧四者之间的关系如图1-1所示，是一个逐步提炼的过程，通过对数据的认知和解读，数据可以转化为信息；通过对大量信息的体验和学习，并从中提取关于事物的正确理解和对现实世界的合理解释，信息可以转化为知识；通过对知识的整合运用，知识可以转化为智慧。数据→信息→知识→智慧→推动人类社会进一步向前发展，数据是这一转变过程中的基础。

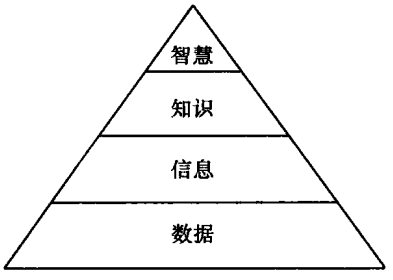


图1-1　数据、信息、知识、智慧四者的关系

## 1.2　数据工程概述

### 1.2.1　数据工程产生的背景

数据工程是信息技术发展的产物，其产生的背景主要有以下四个方面。

**1. 数据资源的开发和利用成为推动社会发展和进步的重要力量**

由于信息技术的发展使数据的资源价值更易发挥，数据的资源特征日益显著。20世纪五六十年代，由于计算技术的发展和成熟，使得大量数据的收集、加工、存储和利用成为可能，致使数据成为可能产生经济和社会效益的重要资源；70年代以来，计算机软件和硬件技术的发展，使对大量数据的精细加工和数据变成信息并加以利用成为可能；当前，由于计算机技术和通信技术的发展，使信息的传播和利用超脱了时空的限制，成为社会发展和进步的极为重要的、可共享的资源，对数据资源的开发和利用成为推动社会进步的重要力量。世界各国非常重视数据资源的开发和利用，比如美国通过数据的流动和应用激励美国经济的发展，确保美国在21世纪信息时代处于世界领先地位。美国国家科学基金会NSF(National Science Foundation Cyber Infrastructure Council)在《21世纪科学研究的信息化基础设施》(Cyber Infrastructure Vision for 21st Century Discovery)报告中明确指出“在未来，美国科学和工程上的国际领先地位将越来越取决于在数字化科学数据的优势上，取决于通过成熟的数据挖掘、集成、分析与可视化工具将其转换为信息和知识的能力。”2006年，联合国成立促进发展中国家科学数据共享与应用全球联盟GAIN(Global Alliance for

Enhancing Access to and Application of Scientific Data in Developing Countries)，强调将科学数据作为协助发展中国家建设的战略资源，实现联合国“千禧年发展目标”(Millennium Development Goal)。

**2. 数据总量呈指数式增长，数据管理困难**

随着信息技术的进步，人类对客观事物的认识不断加深，数据量的规模空前增大。根据IDC所进行的研究计划“Digital Universe”的分析报告，在整个2007年，我们这个世界生成、占用的数字信息及复制总量大约是281EB[①]，这个数据平摊到地球上的所有人，大约是每个人45GB的数据，2008年生成了大约400EB的数据，到2011年，每年产生的数字信息大约是1800EB，10倍于2006年产生的信息量。展望未来，数据有望每18个月翻一番。到2012年创造的数字信息将达到2008年的5倍。全球数据增长趋势如图1-2所示。数据总量呈指数式增长，为数据管理带来了新的挑战。

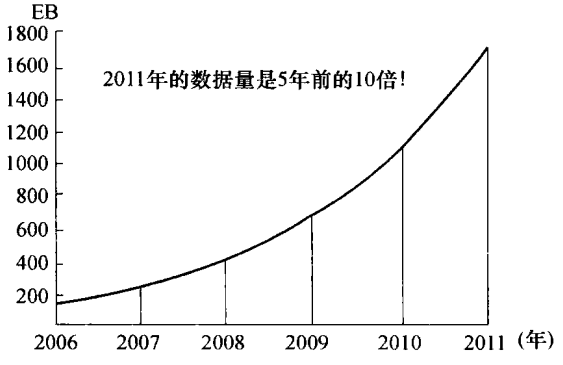


图1-2　全球数据增长趋势

**3. 数据集成与共享的迫切需求**

在20世纪60年代到70年代期间，信息系统应用的主要目标是利用计算机来代替一部分联系不那么密切、手工的重复性劳动的工作环节，以提高生产或管理效率，这一阶段还没有数据集成与共享的需求。

到了20世纪八九十年代，各行业在信息系统上进行了巨大的投资，以满足业务处理和管理需要为目标，建立了众多的应用信息系统。由于各个机构是按照职能来组织各个部门，不同的部门使用不同的应用信息系统来协助他们完成规定的职能，导致许多关键的数据被封闭在相互独立的系统中，形成一个个所谓的“信息孤岛”。

到21世纪，信息技术得到了迅猛的发展，随着各行业的信息化建设向广度和深度扩展，业务需求也在不断变化，需要将众多的“信息孤岛”集成和整合为一个有机整体，实现数据的无缝流动和共享。根据META Group的统计，一家典型的大型企业平均拥有49个应用系统，33%的IT预算是花在信息系统集成上。可以说信息系统集成是各个行业在信息化建设中不可缺少的环节，而信息系统集成的核心是数据集成和共享。

**4. 数据资源建设成为制约信息系统效能发挥的瓶颈**

尽管数据总量在不断增加，但是人们在需要应用数据解决实际问题时，却缺少有效数

① 1EB=1024PB，1PB=1024TB，1TB=1024GB，1GB=1024MB，1MB=1024kB，1kB=1024Bytes。

据的支撑。需要花费大量的人力和财力，采取各种手段，千方百计地去抽取、转换和整合数据。在应用数据时，面临的具体问题主要有：不知道哪里有所需要的数据；知道数据存放的位置，但由于技术或组织的原因无法访问数据；知道数据存放的位置，也可以访问，但是缺少语义信息而无法理解数据；数据可以访问，也可以理解，但是同样的信息，在不同的位置，名称、格式、含义却不同。这些问题的存在，导致信息系统缺乏有效数据的支撑，不能发挥信息系统应有的效能，因此数据资源建设成为制约信息系统效能发挥的瓶颈。

数据工程就是在以上这些背景下产生的一门新兴学科。

### 1.2.2 数据工程的内涵

**1. 数据工程概念**

数据工程是以数据作为研究对象、以数据活动为研究内容，以实现数据重用、共享与应用为目标的科学。

从应用的观点出发，数据工程是关于数据生产和数据使用的信息系统工程。数据的生产者将经过规范化处理的、语义清晰的数据提供给数据应用者使用。

从生命周期的观点出发，数据工程是关于数据定义、标准化、采集、处理、运用、共享与重用、存储和容灾备份的信息系统工程，强调对数据的全寿命管理。

从学科发展角度看，数据工程是设计和实现数据库系统及数据库应用系统的理论、方法和技术，是研究结构化数据表示、数据管理和数据应用的一门学科。

**2. 数据工程研究的内容**

数据工程研究的主要内容包括数据建模、数据标准化、数据管理、数据应用和数据安全等。

（1）数据建模。现实世界中的数据描述现实世界中的一些事物的某些方面的特征及其相互联系，是原始的、非规范化的。通过数据建模，对现实世界中具体的人、物、活动、概念进行抽象、表示和处理，变成计算机可处理的数据，也就是把现实世界中的数据抽象到信息世界和计算机世界。数据建模主要研究如何运用关系数据库设计理论，利用数据建模工具，建立既能正确反映客观世界，又便于计算机处理的数据模型。

（2）数据标准化。数据标准化主要为复杂的信息表达、分类和定位建立相应的原则和规范，使其简单化、结构化和标准化，从而实现信息的可理解、可比较和可共享，为信息在异构系统之间实现语义互操作提供基础支撑。

数据标准化主要是在现有国家、部门、地方和企业的现有标准规范基础上，结合国际相关标准，制定数据标准，并在信息化建设中宣传、贯彻和执行。数据标准化重点研究数据标准化的组成和方法等内容。

（3）数据管理。数据管理是保证数据有效性的前提。首先要通过合理、安全、有效的方式将数据保存到数据存储介质上，实现数据的长期保存；然后对数据进行维护管理，提高数据的质量。数据管理研究的主要内容包括数据存储、备份与容灾的技术和方法，以及数据质量因素、数据质量评价方法和数据清理方法。

（4）数据应用。数据资源只有得到应用才能实现自身价值，数据应用需要通过数据集成、数据挖掘、数据服务、数据可视化、信息检索等手段，将数据转为信息或知识，辅助人们进行决策。数据应用研究的主要内容包括数据集成、数据挖掘、数据服务、数据可视化

和信息检索的相关技术和方法。

(5) 数据安全。数据是脆弱的,它可能被无意识或有意识地破坏、修改,需要采用一定的数据安全措施,确保合法的用户采用正确的方式、在正确的时间对相应的数据进行正确的操作,确保数据的机密性、完整性、可用性和合法使用。

**3. 数据工程与数据库技术之间的关系**

数据库技术主要涵盖数据库原理、结构化查询语言、数据库设计和数据库管理系统(SQL Server、Oracle、DB2、…)等内容,研究和解决计算机信息处理中大量数据有效组织和存储的问题,包括在数据库系统中减少数据存储冗余、实现数据共享、保障数据安全以及高效地检索数据和处理数据。数据库技术是计算机数据处理与信息管理系统的核心,是数据工程的理论与技术基础。数据工程以数据的生命周期为主线,研究数据生命周期中的数据活动,包括数据的定义、管理和应用中的理论、技术和方法。

### 1.2.3 数据工程研究的对象

数据工程研究的对象主要是具有主题的、可标识的、能被计算机处理的数据集合,即数据集,主要指关系型数据库和文件系统,也可以是图像、音频、视频、软件等。数据集的具体含义如下:

(1) 主题:围绕着某一项特定任务或活动进行数据规划和设计时,对其内容进行的系统归纳和描述。将相同属性的主题归并在一起形成相同的类,将不同属性的主题区分开形成不同的类;主题还可被划分成若干子主题或子子主题。

(2) 可标识:指能通过规范的名称和标识符等对数据集进行标记,以供识别。标识与名称的取值需要通过具体的命名或编码规则来规范。

(3) 能被计算机处理:指可以通过计算机技术(软硬件、网络)对数据集内容进行发布、交换、管理和查询应用。这些数据可以由不同的物理存储格式来实现,按照数据元的定义与数据类型,在计算机系统中以数值、日期、字符、图像等不同的类型表示。

(4) 数据集合:指由按照数据元所形成的若干数据记录所构成的集合。例如,学生管理数据集由学生基本信息、学生工作简历信息、政治面貌信息、家庭成员信息等不同数据组成。

## 1.3 数据工程的现状与发展

近些年来,数据工程在很多领域有了较大发展,尤其在军事领域,以美军为代表的发达国家为了建立数据共享基础设施,持续有效地推动了数据工程的发展。

### 1.3.1 美军数据管理策略的演进

在军事领域,美军的数据工程建设代表了该领域发展水平。美国国防部 DoD(Department of Defense)于第二次世界大战之后着手进行国防物质编目工作,并且于 20 世纪 90 年代启动了数据工程(Data Engineering),实现了国防数据词典系统 DDDS(DoD Data Dictionary System)、数据共享环境 SHADE(Shared Data Engineering)和联合公共数据库 JCDB(Joint Common Database)。这些成果为实现 $C^4ISR$(Command Control Communication

Computers Intelligence Surveillance and Reconnaissance)系统之间的数据重用和数据共享奠定了基础,也确保了美军在海湾战争、科索沃战争、阿富汗战争和伊拉克战争中的信息优势。2003 年美军在伊拉克战争中主要依靠 DDDS、SHADE 和 JCDB 等技术手段实现了 95% 以上的信息共享。

数据管理策略的演进在一定程度上反映了人们对数据工程的认识和实践在不断地深入。以美军为例,迄今为止,美军数据管理策略的演进过程从总体上经历了三个阶段。

**1. 统一的数据管理阶段**

美军为使指挥控制、后勤、情报、人事和财务管理等自动数据处理系统之间进行数据交换并提高系统间的兼容性,从 1964 年开始对数据元素和代码实施集中统一的管理。1964 年 12 月 7 日,美国国防部颁布了国防部指令 DoDD 5000.11《数据元素和数据代码标准化大纲》,随后,陆续制定了一系列配套文件,以实现数据元素和代码标准化目标。在 DoDD 5000.11 系列文件的指导下,美军各军种也制定了相应的数据管理文件来贯彻国防部的数据管理要求。标准数据元素和代码在采办、后勤、指挥控制等许多领域得到了比较广泛的应用,不仅促进了数据系统之间的数据交换,改善了系统之间的兼容性,还有效地提高了数据采集和数据处理的效率,减少了数据的冗余和不一致性。该阶段的数据管理呈现如下特点。

(1) 明确了统一的管理流程。该阶段的管理流程是,将所有标准数据元素都收入到由国防部统一管理的一个集中的主文件(A Centralized Master File)中,通过分发和更新这个文件,各有关部门即可查阅到需要的数据元素。如果其中没有能够满足需要的标准数据元素或原有的标准数据元素需要修改,则各部门根据业务需要再提出新的候选标准数据元素或修订方案并提交到国防部组织审查,审查通过后即扩充到标准数据元素集中。

(2) 建立了两个层次的管理机构。在统一的数据管理阶段,美国国防部的数据管理机构主要分为国防部和国防部的各部局两个层次。国防部主管审计的国防部长助理负责制定数据管理政策、规程等;审查、批准和颁布标准数据元素及代码;监督检查各军种使用标准数据元素的情况;协调并解决与数据管理有关的问题。各军种/部/局分别负责与其自身业务密切相关的数据元素的标准化(通用基础的数据元素由国防部负责),其主要工作是在相应工作组的配合下,对数据元素进行标识、定义、分类和编码,并作为候选的标准数据元素提交国防部审查;对其他部局提交的标准数据元素提出修改意见;在系统建设过程中积极贯彻实施国防部已颁布的标准数据元素。

(3) 数据标准化工作存在一定的局限性。国防部统一管理所有的标准数据元素和标准代码,制定管理流程和管理机构,主要以文件形式进行统一管理。此时的数据标准化存在一定的局限性,主要在于统一数据元素的名称、缩写、定义、值域及值域代码。没有相应数据管理支撑工具,对数据标准的修订和使用情况跟踪均带来问题。

**2. 集中的数据管理阶段**

随着信息技术的发展,美军各军种及功能域(如指控、后勤、财务、医疗等)都拥有了数量可观的信息系统,此外诸多的武器系统也不同程度地实现了信息化,这使得美国国防部不得不面对两方面的挑战。一是互操作性带来的挑战,即如何使这些系统之间更有效地交换信息,如何实现不同系统之间的数据共享;二是数据标准化带来的挑战。美军在 DoDD 5000.11 的指导下虽然已对许多数据进行了标准化,但是其深度(主要局限于数据