

# 建立水土保持数据库的探讨

李 壁 成

(中国科学院西北水土保持研究所)

当前,我国正面临着世界新技术革命的挑战,中央领导同志和一些著名科学家都希望我们赶上这个浪潮,把我国四化建设很快搞上去。在这场新的技术革命中,数据库技术异军突起,引人注目。这是因为与物质、能源一样,信息也是重要的资源,而数据库系统是信息处理系统的核心,已广泛应用于科学研究、工业控制、军事情报、经济和资源管理等各个领域,成为现代信息社会的重要工具。因此,研究和应用数据库技术,是水土保持学科迎接新技术革命挑战的最有效途径之一。

## 一、什么是数据库

人类活动的整个历史,一刻也离不开对信息和数据的收集、保存、利用和处理。原始社会人类仅能用语言、绘画、火光等传递信息,当发明文字、纸张和印刷技术以后,就以纸张为介质,通过文字记载、收集和保存信息。近代电子技术出现后,人们又利用磁性材料作介质来收集和保存信息,使得信息利用和加工进入了更高级的阶段。

本世纪六十年代以来,社会生产力高速发展,新技术层出不穷,信息量急剧膨胀,使整个人类社会正在迈向信息化的社会。人们对信息和数据的利用和处理已进入全面自动化、网络化和国际化的新阶段。例如,查找浩如烟海的情报资料、档案,处理繁杂的银行帐目、资金往来,进行庞大的人口统计分析,以及国民经济建设和科学研究等,这些工作既需要利用大量数据,又要求快速处理,及时得到结果。如果仅由人工查找和处理,是难以完成的,甚至是不可能的,所以迫切需要借助电子计算机的高速度和大容量来处理信息和数据。为此,数据库作为计算机软件的一个重要分支,近十几年才迅速崛起。

“数据库”一语的来源目前难以确定,但众所公认C. W. Bachman是数据库领域的早期先锋,他在1962年设计了集中数据库(IDS)系统,并于1963年投入运行。数据库的历史虽短,但其发展速度却十分惊人。据估计,目前国际上生产数据库的厂家达200多个,其中59个既是生产者又是服务者。全球已拥有近万个大型数据库。其应用已深入到人类生活的各个领域,如搜集、检索情报资料,管理档案,制定国民经济计划,企业管理,银行业务,交通运输等等,并成为计算机科学其它领域(如人工智能、计算机辅助设计等)必不可少的有力工具,甚至渗透到人们的日常生活和学习中。如美国最近出现的为公众服务的“信息银行”,人们从中可检索到各自需要的新闻时事、天气预报、文娱节目、电话号码、学习指南等各种信息。总之,数据库日益成为人类

活动不可缺少的公共资源，正在逐步发展成为新的产业——第四产业（信息服务）。

我国数据库技术起步较晚，自1977年召开数据库学术会议以来，也有了较快的发展。目前，林业部建立了森林资源数据库系统、中国科学院的科学数据库、国家经济信息系统数据库等一批大型数据库，已列入重点工程计划，准备建库。

不难理解，数据库系统是实有有组织地、动态地存储大量关联数据，方便多用户访问的计算机软硬件资源组成的系统。它与文件系统的重要区别是数据的充分共享、交叉访问与应用（程序）的高度独立性。其特点是：有数据结构，能建立型与型之间的关系；提供数据共享，追求小的冗余度；物理模块和逻辑的独立性，因而可以对数据进行收集、加工、管理、传输等一系列活动，使本来一些杂乱无章的数据产生巨大的价值。

一般来说，数据库系统由数据库（DB）、数据库管理系统（DBMS）和数据库管理员（DBA）三部分构成，其体系结构如图1。

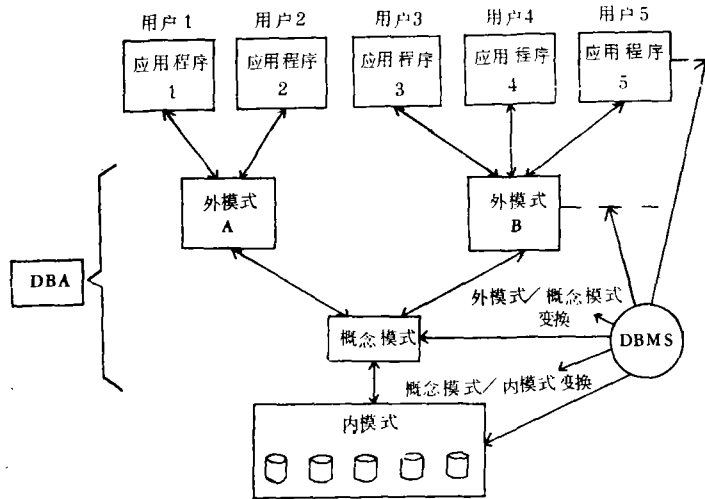


图1 数据库系统的体系结构

数据库的核心问题是数据模型，它是数据库中数据的整体逻辑结构，反映实体内部和实体间的联系。如果用数学公式描述，数据模型DM可以定义为： $DM = \{ R, L \}$ 。式中R是纪录型的集合，L是记录型间联系的集合。目前流行的数据模型有关系模型、网状模型和层状模型三种。在空间信息系统中，基于属性数据模型应用较多。当今，数据库的应用愈来愈广泛，要求越来越高，推动着数据库的不断发展。八十年代数据库技术主要研究方向是分布式数据库系统、数据库机、专家系统、用户接口以及基础理论研究等。随着上述研究工作的突破，一种使用方便、通用高效的分布式数据库和具有人工智能的专家系统将展现在人们面前，使人类的信息数据处理进入一个崭新的时代。

## 二、水土保持与资源数据库的关系

水土保持是根据土壤侵蚀规律，采取综合防治措施，实现合理利用水土资源的科学。它与资源数据库有着密切的联系。所谓资源数据库，就是在空间技术和环境信息系统发展的基础上而产生的以资源数据为登记项的数据库。其特征是将资源信息按照一定的空间地理座标，以统一格式

输入计算机建库，供用户查询、检索使用，具有空间分布性质和图像处理等功能，其系统结构见图 2。

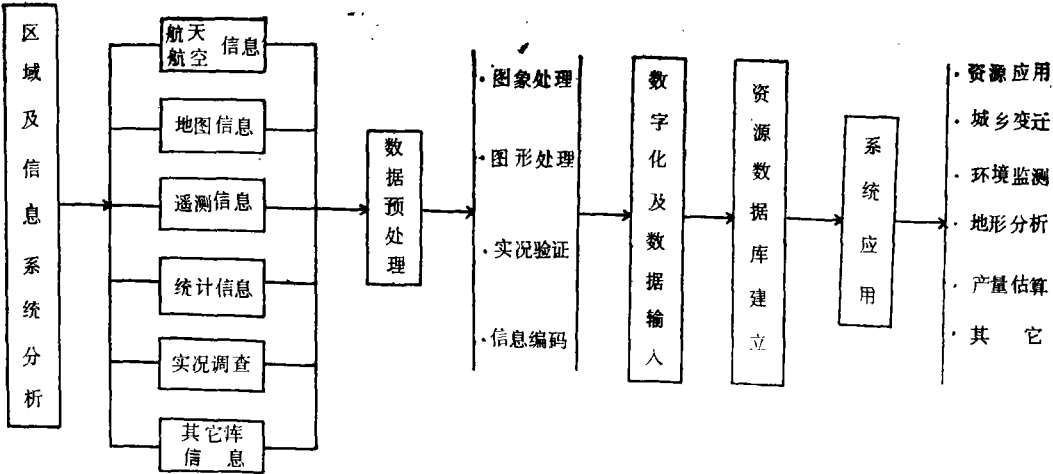


图 2 环境信息系统与资源数据库结构图

由于空间技术的飞速发展，环境信息与日俱增，这为资源数据库提供了丰富的信息源，从而

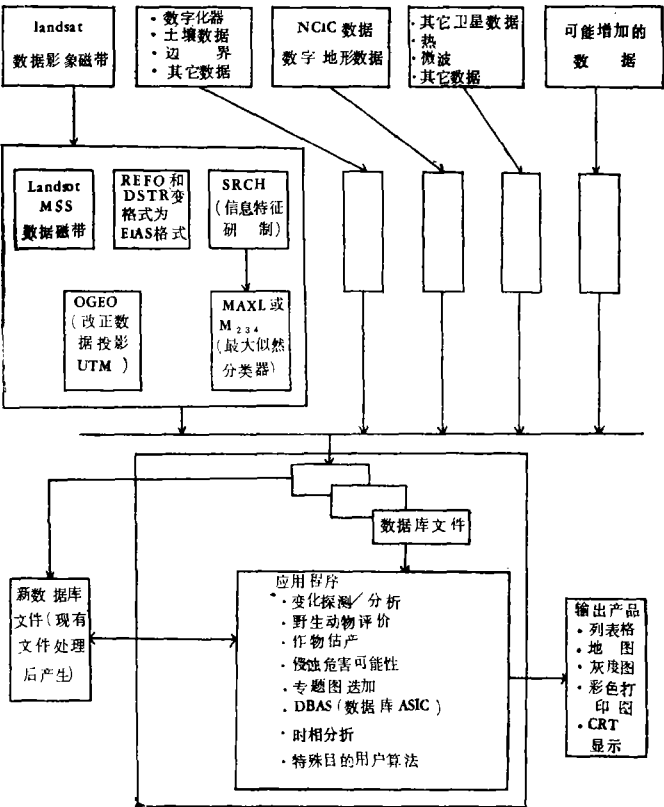


图 3 ELAS地理信息系统

推动了资源数据库的发展。美国、日本、加拿大、瑞典等国都建立了各种资源数据库。日本从1968年开始，在全国建立了三级格网，最小一级格网为1平方公里，包括16个数据；三级格网可以互相转换，内容包括土地利用规划、交通噪声、垃圾处理、水资源等，可为50多个部门服务；加拿大则按多边形结构建立了土地资源数据库；美国国家航空和航天管理局（NASA）开发的ELAS地理信息系统（见图 3），不仅用途广泛，而且也为水土保持数据库提供了依据。

国内不少单位也开始了这方面的研究和应用，如中国科学院遥感应用研究所在四川盐边县进行的区域信息系统和资源数据库的开发研究工作等。这些都为建立水土保持数据库提供了宝贵的经验。认真学习国内、外数据库的

经验，深入研究我国水土保持信息和数据结构，定能加快开发和建立水土保持数据库的进程。

### 三、建立水土保持数据库的设想

我国水土保持有着悠久的历史，在世界上也有很大影响，特别是建国三十多年来，积累了大量信息资料和数据，因此建立水土保持数据库不仅十分必要，而且也是完全可能的。但是一个大型、功能良好的数据库的建立和维护，是项艰巨工程，它需要数据库管理员、专业技术人员、计算机技术人员、数据库应用人员等多方面密切配合，以及技术上、物质上的互相合作。建库必须从价值、效益出发，综合考虑系统的完整性和可扩充性，尽可能地提高存取速度，这就需要有组织、有计划、有步骤地开展工作。

根据美国国际商业机器（IBM）公司研究，数据库的建立有六个阶段，即系统需求分析、系统设计、数据库建立实施、应用开发研究、系统运行和维护以及系统评价。其中系统需求分析是基础，它不仅需要充分了解用户的各种要求，而且要对大量复杂数据进行分析处理、摸清数据的有效性、完整性、冗余性、权威性、标准化、表示方法、总量、来源等，从而采取有效、灵活、经济地数据处理办法。这一阶段所用时间一般约占到全部时间的60%。而逻辑设计又是数据库的关键，因为逻辑结构的设计是对数据库全体数据的综合分析和合理布局，其结构的好坏直接影响

到数据库的效率、存取速度、灵活性、经济性和可靠性等指标。一个数据库如果没有一个坚实可靠的数据逻辑结构作基础，将会使今后的实际应用造成许多漏洞，甚至不得不推倒重来，从而造成很大的浪费。

当然，数据库系统也有大和小、复杂和简单之分。环境信息系统就全国而言，可分为三大类，即全国性综合系统、区域信息系统和专题信息系统。因此，水土保持数据库也可考虑先易后难，由小到大，从专业到综合，由区域（自然区或行政区）到全国，逐步逐级建立数据库，探索发展，在实践中不断前进。

如何建立水土保持数据库呢？根据数据库原理和水土保持数据结构的实际情况，我初步设想要抓好四个环节，即数据的搜集和系统分析，逻辑数学模型设计，物理设计和运用设计。其它步骤也不能忽视（见图4）。

**1、数据的搜集和系统分析。**数据库管理的对象是数据，因此搜集有关数据是建库的前提条件。数据库应包括哪些数据、数据量的大小、产生哪些报表等，都必须调查清楚，搜集齐全。数据的搜集工作主要是在系统调查和系统分析过程中进行。

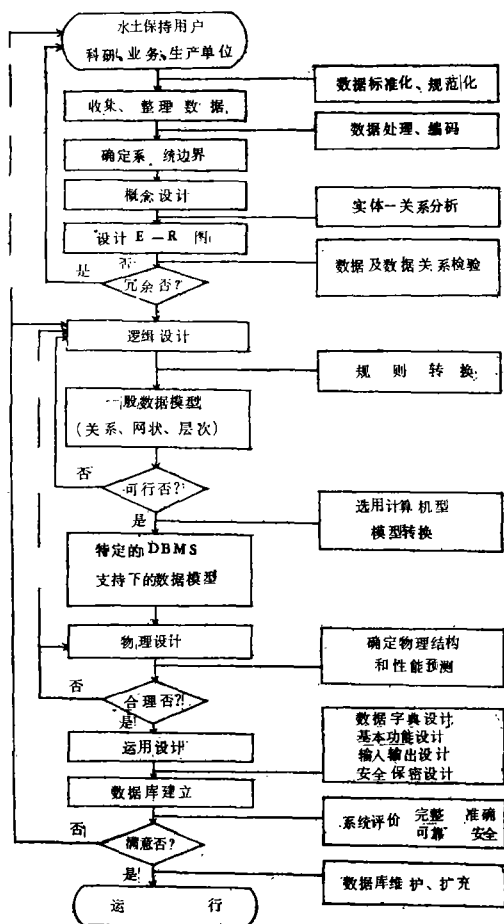


图4 水土保持数据库系统建立工作流程图

这是一个反复调查，逐步深入的过程。目前水土保持专业术语比较混杂，有关数据比较零乱，为此首先要进行数据的标准化和规范化，这需要作大量艰苦细致的工作。

同时还应确定系统边界。一是要确定水土保持系统的学科边界，因水土保持科学是综合性很强的学科，但它也不是包罗万象的大杂烩，必须分清哪些是学科本身范畴，哪些是相邻或渗透学科，只有这样才能确定数据的范围和数据间的关系；二是要确定计算机系统的范围和它同外部环境之间的相互关系，即哪些活动由计算机去执行，哪些仍由人工完成，根据系统环境和条件，决定系统功能。

在完成数据搜集和系统分析的基础上，即可进行概念结构设计。这种设计的主要特点是能充分反映现实世界，易于理解，可以和不熟悉计算机的用户交换意见，当现实环境改变时容易修改和扩充，并易于向关系、网状、层状模型转换。因此概念结构是各种数据模型的共同基础。描述概念结构的有力工具是“E—R图”（实体—关系图），它包括实体、关系和属性三个基本成分。如水土保持的实体是自然界客观存在的“水”和“土”，其关系是指水土与自然界的联系——“保持”，属性为“某地域的水土”，如“黄土高原的水土”。在初步的“E—R图”中可能存在冗余的数据和实体间冗余的联系。所谓冗余的数据是指可以由基本数据导出的数据，冗余的联系是可以由基本联系导出的别的联系。冗余信息的存在会破坏数据库的完整性，给数据库管理增加很多困难，应尽可能加以消除。在消除冗余而形成“基本E—R图”后，还应反馈回去，征求用户意见，这样反复几次最终形成“基本E—R图”。

**2、逻辑数据模型设计。**逻辑设计的任务，是把概念结构转换为DBMS（数据库管理系统）所支持的数据模型相符合的过程。设计逻辑数据模型应选择最好的数据模型和最合适的DBMS，但实际情况往往由于计算机已经确定，设计人员没有多大选择余地。现行的DBMS一般只支持关系、网状、层状三种模型的一种，对一种数据模型，各机器系统又有许多不同的限制。因此“E—R图”向逻辑模型转换时，必须熟悉DBMS的性能和要求。规范化理论是判断关系模型优劣的标准，也是模型转换的指南和工具，一般均先将概念结构模型向关系模型转换，然后将关系模型向DBMS所支持的格式化模型转换。

对于水土保持数据库来说，逻辑数据模型的转换和设计，是建库的成败关键之一。因为水土保持应用计算机较少，究竟哪种机型适宜建立DBMS，还需要摸索和开发软件，这一工作量也是很大的。因此建立数据逻辑结构和模型，需要全面地、周密地分析和反复试验，从而选择比较合理的逻辑数据结构和DBMS

**3、物理设计。**逻辑数据模型确定后，选定一个最合适的应用环境的物理结构，就是物理设计的任务。物理结构主要是数据库的存储结构和存取方法，它依赖于计算机系统。其内容包括确定数据存储结构，存储路径的选择和调整，即数据库必须支持多个用户的多种应用，因而必须提供数据库的多个存取入口或多条存储路径。如水土保持数据库既要考虑为业务部门存储和提供决策数据，又要考虑科研机构存储和提供科学实验数据，还要考虑生产单位的生态经济数据等，这就需要设计者根据建库目的和用途，进行定量分析计算，确定存取路径，确定数据存放位置。如把易变数据与稳定部分分开，把经常存取与不经常存取的数据分开，将经常存取或时间要求快的记录放在高速存储器上，其它放在低速存储器上等，确定数据块大小。由于系统存取数据是以块为单位，而应用程序存取数据是以记录为单位，为了充分利用存储空间和提高存取效率，数据块大小必须与记录长度相匹配，确定缓冲区大小。应用程序读取数据和存储数据都要通过缓冲区，因而缓冲区大小直接影响存取效率和存取空间的利用，所以应建立适度的缓冲区。

物理设计过程需要对时间、空间效率，维护代价和各用户要求等因素进行权衡折衷，其结

果会产生多种方案，这就需要进行综合评价和性能预测，从中选优。

**4、运用设计。**这是系统设计的最后一道工序。主要包括数据字典、基本功能、输入输出和保密安全等设计。在数据库中，由于数据量大，关系复杂，多用户共享等情况，数据字典（Data Dictionary/Directory，记为DD/D）显得尤为重要。它是数据处理人员在数据库的设计、实现、运行、维护及扩充各阶段中控制并管理有关数据的信息工具。DD/D包含每一数据元的名字、意义、描述、来源、职责、格式、用途以及数据间的关系等等，因此DD/D可称为数据的数（元数据）、特殊的数据库和信息系统的核心。它与DBMS同为数据库管理中的两个主要工具，二者在数据资源的管理与使用中互相补充。同时DD/D用途也很广泛，它作为程序自动生成的工具已引起人们的研究兴趣。为此在水土保持数据库的研制中，设计一个功能良好的DD/D，是值得研究的重要课题。

在全面完成上述工作后，就要进入数据库的建立实施阶段。系统实施仍是整个系统设计的继续。系统设计虽然是在认真进行系统分析的基础上进行的，但是数据库系统是大型软件工程，各部分之间互相联系、互相制约，任何差错都会影响整个系统的可靠性。因此在实施阶段还要检验设计、反馈信息，进一步改进系统设计和结构。系统实施的主要内容是数据库模拟、数据库试建、创造数据库、数据入库等。经过这一系列复杂的工序，一个新的数据库系统就诞生了。

但数据库建立后还要进行最后一项重要工作，就是系统评价。这也是一项相当复杂的工作。因为一个数据库的成功与否，要从多方面衡量，涉及许多因素，而且新系统的功能和效益也不是一下全部显示出来，加上各用户又有不同的要求，因此对系统应作具体而又合理的分析评价。一般着重评价系统的合理性和实用性，可靠性和完整性，存取效率，经济和社会效益以及学术价值等。通过系统评价和签定后，数据库即可投入运行，为用户提供服务。在系统实际使用中，一方面还会出现设计和实现中的问题，必须不断排除；另一方面系统的功能还应逐步扩充和完善。

数据库系统是一个新兴研究领域，也是各个学科实现信息化的有力工具。由于我们对数据库和水土保持专业知识都有限，上述肤浅设想很不成熟，期望更多的科技人员投入这方面的研究。我们相信，随着数据库技术的深入发展，在广大科技人员的努力下，水土保持数据库，甚至水土保持知识库和专家系统，将会逐步建立起来，在四化建设中发挥巨大经济效益，使水土保持科学发展进入一个崭新的阶段。