

长江流域梅雨锋暴雨外场科学试验资料库系统设计与实施*

高梅 接连淑 张文华 贾鹏群 仪清菊

(中国气象科学研究院,北京 100081)

2001-2002年在长江中下游地区开展了梅雨锋强暴雨的外场科学试验。试验期间收集和了大量的原始观测数据和分析加工产品。日本科学家携带部分观测设备参加了试验,获取的观测资料也一并纳入数据库中。数据共分为12大类,约800GB。目前,该资料库已正式在中国气象科学研究院的Internet网站上运行,用户可以通过登录气科院网址访问该资料库。

1 资料库设计思想

长江流域梅雨锋暴雨外场科学试验资料库的设计与建设是以长江中下游地区开展的梅雨锋强暴雨外场科学试验为数据基础。其基本原则是满足用户的最大需求,实现数据共享。与传统数据库的最大区别在于该资料库的建设不仅要达到管理数据的目的,更重要的是能够提供基于网络的数据服务和共享。因此,系统整体设计包括两个部分:数据库系统设计和信息共享系统设计。最终要形成一个架构在Internet上的、能够提供基于Web浏览器动态访问的、由一系列气象数据库群组成的气象科学试验数据共享系统。图1是基于网络的长江流域梅雨锋暴雨外场科学试验资料库系统结构示意图。

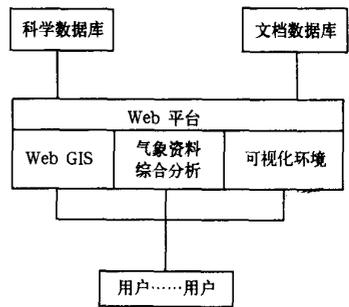


图1 长江流域梅雨锋暴雨外场科学试验资料库系统体系结构

图中资料库系统包括了数据层、应用层和会话层。数据层提供了两类不同属性的数据源:科学数据和文献数据。应用层通过各种可视化显示平台、分析工具和数据查询显示界面,为用户提供方便、实用的数据共享环境。最终通过网络为用户提供各种数据共享服务。

2 数据库系统设计

在外场试验中,对暴雨过程的加密观测和数据收集(见表1)是试验的主要目的。因此,暴雨个例观测数据是资料库数据收集的重点。建议重点研究的暴雨个例过程有:

2001-06-16 ~ 2001-06-19, 2001-07-11 ~ 2001-07-15, 2001-08-05 ~ 2001-08-09
2002-06-21 ~ 2002-06-24, 2002-07-22 ~ 2002-07-23。

本文资料库的数据管理系统设计原则为:

(1) 面向用户需求,提供典型暴雨个例数据集 为了提高数据库的使用效率,数据库设计本着以分类数据收集为基础,以暴雨个例数据综合为核心,提供最有研究价值的、资

* 本研究得到“973”中国暴雨研究项目(G1998040900)的资助。
2003-09-15 收到,2003-11-04 收到修改稿。

表 1 长江流域梅雨锋暴雨外场试验的资料收集

序号	1	2	3	4	5	6	7	8	9	10	11	12
种类	地面	探空	降水	闪电	风廓 线仪	雷达	卫星	边界 层	GPS	格点 分析	日方 交换	自动 站
数量 (GB)	0.27	0.3	0.03	0.02	11	314	317	5.7	0.002	7.8	140	0.6
合计	约 807GB											

料收集最完整的、最具有代表性的个例分析资料集。使用户从数据库中直接可以得到与某个个例相关的完整的数据资料。

(2) 分类建库,综合检索 本数据库管理的数据分成两大类:气象科学数据和文档数据。面向不同的数据对象分别建立相应的数据组织结构,生成相应的数据库模型,保证数据管理的合理性。同时通过建立相应的视图和查询,实现不同数据间的相互关联和查询。

(3) 针对不同属性的数据,采用不同的数据管理方式 站点观测数据全部进入数据库,按照数据的观测方法或处理方法进行分类管理。格点场数据或图形图像等二维数据,采用数据库管理和文件管理相结合的管理方式。数据库管理系统只管理到数据存放的物理位置和路径。

(4) 引进元数据管理模式,加强数据的可读性和可用性 建立关于数据描述和说明的元数据库,用户通过浏览元数据,即可了解该数据集的基本情况,提高了数据的可用性和可检索性。

(5) 建立数据间的相关性,实现联想式文献检索和关联式数据查询 建立不同类型数据间的相关性,使用户查询某类数据时,可以得到相关的其他数据或文档的提示。如:用户查询某时次的雷达资料时,系统会自动列出与该时次雷达数据相关的探空、闪电、GPS 数据以及相关的文档。提高了数据的检索效率。

(6) 设计多种数据访问和数据检索方式 本数据库设计了数据集分类查询和精确查询两种获取数据的方式。分类查询可以得到某个数据集的完整资料,精确查询可以得到某类数据某个时次、站点、天气要素的精确的观测数据。对于雷达数据提供了通过检索索引图得到相关数据的数据检索方式,使用户可以直接浏览所需数据的情况。用户访问资料库获取数据的流程如图 2 所示。

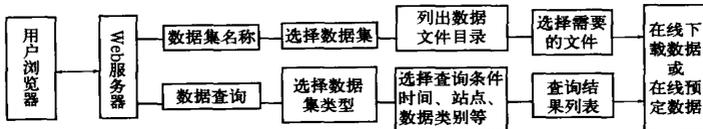


图 2 用户访问资料库获取数据的流程

(7) 基于 WebGIS 平台,提供可视化的用户信息查询界面 将资料库的查询显示与地理信息系统相结合,建立外场试验空间信息查询系统,使数据使用者在可视化的环境下动态了解数据产生的背景,直观得到与地理信息相关的数据。特别是试验区各雨量站的逐时雨量值及逐时降水分布图等。

(8) 综合性 本数据库全面收集和提供了两年外场科学试验获得的各类观测数据、

相关文献和研究开发成果。形成了从资料收集、处理、管理、成果集成到提供基于网络的数据共享服务的组织管理体系。综合性地提供了外场科学试验丰富的信息资源。

(9) 数据的预处理和质量控制 从现场观测入手,把好数据的质量关。实行测站、市局和省局三级质量控制,专门编制了加密资料审核程序,采取计算机审核和人工辅助审核相结合的方式对数据的质量把关。项目资料组对收集到的原始观测数据进行译码、格式检查、格式转换和标准化,统一文件命名规则,对数据进行极值检测和面质量控制,这些措施确保了数据的质量和规范性。

3 信息共享系统设计

信息共享系统设计时主要考虑以下原则:

(1) 提供多级用户身份管理 系统安全是信息共享系统设计的一个关键。通过对信息的分级分类,实现对用户访问权限的限制和管理。本系统将数据分成内部共享、有条件共享和完全共享三类。在数据分类基础上建立三级用户访问身份:内部用户、有条件外部用户和普通用户。用户身份认证体系,提供了用户身份认证和基于组级的身份访问控制,以便确定不同级别的用户能够访问的资源和应用。数据类别、用户级别和访问权限间的对应关系如表 2 所示。

表 2 数据类别、用户级别和访问权限间的对应关系

数据类别	用户级别	访问权限
内部共享	内部用户	共享所有数据,定制所需的资料内容;使用在线资料分析和显示工具,大致检验资料的质量。
有条件共享	有条件外部用户	浏览数据目录,查看和下载有关资料说明,注册后可在线下载制作好的数据。
完全共享	普通用户	浏览数据目录,查看和下载有关资料说明,在线申请定制的数据。

(2) 建立一套适合科学试验数据库管理和服务的数据库规范和标准 规范数据格式,统一文件命名,便于用户查询和使用。

(3) 平台的独立性 所选择的 Web 接口技术应能运行在不同的操作系统平台上。能和系统中现有的 Web 服务器技术无缝结合,实现 Web 服务器与数据服务器之间的动态连接,从而实现从客户端到数据库服务器的动态访问。同时便于系统的移植和扩展。

(4) 提供可视化的交互式的用户操作界面,用户界面友好,信息查询简单、快捷。

(5) 提供网络导航系统和多种用户在线使用工具,提高数据的可用性。

4 资料库系统实现的功能

(1) 提供了基于网络的数据查询显示界面。用户端通过浏览器即可进行数据查询。

(2) 建立了项目宣传网站和数据集成管理平台。为人们了解项目的进展和获取外场试验数据提供了环境和场所。

(3) 提供两种方式的数据访问:数据集查询和数据库查询。

(4) 数据检索结果以多种格式提供:纯文本、二进制、GRIB 码、标准图形图像格式等。

(5) 利用 WebGIS 平台动态展示外场试验站点分布和相关站点的资料状况。

(6) 给出简单检验数据质量的可视化工具。

(7) 实现多种资料提供方式:在线查询和下载;以 E-mail、光盘、磁盘或磁带等方式为资料使用者提供所需要的资料。