

GIS 数据共享机制研究

蔡爱民¹, 查良松²

(1. 滁州学院 地理系, 安徽 滁州 239012; 2. 安徽师范大学 国土资源与旅游学院, 安徽 芜湖 241000)

摘要: 实现 GIS 数据共享已成为 GIS 发展的迫切需要, 分析了需要数据共享的原因, 评述了当前处理数据共享的几个途径, 认为实现 GIS 数据共享的根本途径是制定全球统一的数据格式, 辅之以政府的宏观调控和引导。

关键词: GIS; 数据共享; 数据格式

中图分类号: P91 **文献标识码:** A **文章编号:** 1001- 2443(2005) 02- 0226- 04

近 20 年来, 随着 GIS 的广泛应用, 我国地理信息产业得到了迅猛发展. 为了提高数据获取和生产的效益, 人们不断对现有数据进行二次开发. 各部门在开发地理信息系统时, 通常根据本部门的特定情况采用不同的数据建模方法, 选用不同厂商的 GIS 软件. 由于不同软件间的数据格式差异, 使得在不同 GIS 软件上开发的系统间的数据交换产生困难^[1]. 事实上, 如何实现 GIS 数据共享已经是限制 GIS 发展的瓶颈问题, 严重的阻碍了 GIS 走向大众化、网络化、产业化.

1 GIS 数据共享的提出

实现 GIS 数据共享是 GIS 发展的一个不变目标^[2], 当前已经显得比任何一个时期都重要. 多种格式数据的并存, 是有其客观原因的.

1.1 数据获取手段多源性

获取数据的手段复杂多样, 形成多种格式的原始数据并存的局面, 如图表、遥感、GPS、统计调查、实地勘测等. 这些不同手段获得的数据其存储格式、提取和处理手段都各不相同. 多数据源的存在, 导致了产生多格式 GIS 数据的可能. 操作这些数据源的平台往往比较单一, 需要 GIS 数据共享机制来解决.

1.2 多 GIS 平台的存在

GIS 平台作为采集 GIS 数据的重要手段, 是影响数据共享的主要原因. 近二十年来, 一批自主知识产权的国产 GIS 软件日趋成熟, 与国外知名软件相比, 我国 GIS 软件有一个明显的瓶颈问题——数据接口, 经常出现数据进不来、出不去或转换过程中信息丢失. 而由于每个 GIS 平台的长处不同, 具体应用中常需要将几个平台结合着使用, 使用户处于一种矛盾的境地. 这种境况一方面是因为 GIS 软件厂商对数据接口的重视程度不够, 另一方面是因为缺少统一的数据标准. 加上 GIS 应用系统很长一段时间处于以具体项目为中心孤立发展状态中, 很多 GIS 软件都有自己的数据格式, 这使得 GIS 的数据共享问题变得尤为突出.

2 数据共享机制的实现途径

考虑数据共享问题, 几乎成了开发每一个 GIS 系统前需要首先考虑的问题. 为了解决数据格式差异带来的矛盾, 实现多源数据共享, 当前的处理方式大致有三种, 即: 数据格式转换模式、数据互操作模式和直接数据访问模式^[3].

2.1 数据格式转换模式

这是当前广泛采用的模式, 先期开展的研究也比较多. 在这种模式下, 其他数据格式经专门的数据转换程序进行格式转换后, 复制到当前系统中的数据库或文件中. 几种重要的空间数据格式有: ESRI 公司的

收稿日期: 2004- 02- 17

基金项目: 安徽省自然科学基金重点项目(00021087); 安徽省教育厅自然科学研究重点项目(2002Kj135ZD).

作者简介: 蔡爱民(1978-), 男, 安徽肥东人, 讲师, 硕士, 从事遥感与地理信息系统的教学与科研工作.

Coverage(Arc/ Info)、ShapeFiles(ArcView)、E00 格式; AutoDesk 的 DXF 格式和 DWG 格式; MapInfo 的 MIF 格式; Intergraph 的 dgn 格式等. 数据在不同系统间转换主要有 3 种实现途径: (1) 数据交换程序; (2) 工业实用标准; (3) 国家空间数据交换格式.

2.1.1 基于数据交换程序的数据转换 通过交换程序将独立系统间的专用数据格式进行转入转出是应用最广的数据转换方法. 图 1 显示了该方法的逻辑结构和 GIS 交换程序进行的数据流程.

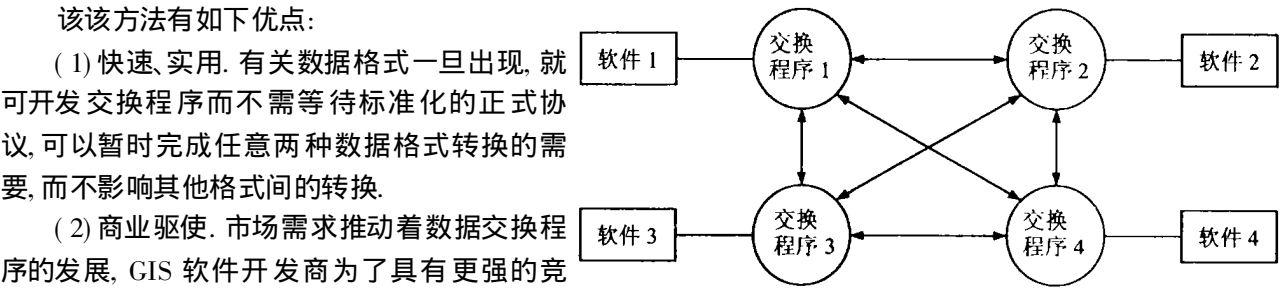


图 1 基于数据交换程度的数据交换格式

(3) 由于只考虑两种数据格式, 很容易开发一种数据交换程序, 逻辑规范和数据模型在只有两个开发商的合作中可得到更好的转换和实施.

该方法的明显缺点是:

(1) 为所有的相关系统开发各自的交换程序, 代价昂贵, 这就使得开发商们为了节约成本而故意忽略一些较少通用的数据格式.

(2) 经常出现软件开发商提供的数据交换程序过时, 或与目标系统现用的格式不兼容.

(3) 有时系统间的逻辑规范和数据模型不兼容, 或对其它数据模型的有限理解, 导致了交换过程中的数据丢失.

(4) 对用户(特别是数据供应者)而言, 提供各种专用格式的数据是困难且不划算的, 数据供应者为了迎合需求常需要重新购置软件.

2.1.2 基于工业实用标准的数据转换 在不太复杂的 GIS 环境中, 将工业实用标准——广泛应用的软件包(例如 DBF 格式)具备的格式选作系统间的通用中介格式, 这种方法被认为是一种可行的临时数据标准化方法^[3]. 通过转入转出该中介格式就完成了数据交换. 例如在两种 GIS 软件 Spatialist 与 MapInfo 之间进行数据转换, 为了保证空间和属性数据完整和准确, 采用 AutoCAD 作为数据交换的平台, 将 AutoCAD 下的图形与 FoxPro 的 DBF 文件属性数据写成 MapInfo 能够接受的 MIF、MID 文件. 然后利用生成的 MIF、MID 文件转入 MapInfo, 形成 MapInfo 的空间和属性数据; 也可利用 MapInfo 转出 MIF、MID 文件生成 AutoCAD 下带扩展实体数据的实体和 FoxPro 的实体属性数据库. 图 2 显示了系统间这种方法的逻辑结构与数据流通.

为了描述工业实用标准, 数据格式需满足以下准则:

(1) 软件包必须具有广泛的设备基础并在相当长的时间内控制整个市场. 但不一定是目前的主导软件, 只要在过去占领过市场就可. 一个典型的例子就是 dBase 的 DBF 格式, dBase 将其建立为基于计算机的 DBMS(数据库管理系统)实用标准, 尽管 dBase 不再是现在市场上 DBMS 软件的主要竞争者, 但 DBF 仍是广泛应用的数据交换格式之一.

(2) 必须是开放性的, 广大用户可以免费或以极低的花费得到该数据模型及格式规范.

(3) 必须简单易行.

(4) 必须是持久的, 持久的格式有足够的时间让工业和市场接受. 格式不固定, 软件开发商就会因其数据交换程序有可能过时而不愿采用该格式作为相互转化的途径.

使用实用标准的优点:

(1) 可行好并可以立即实施.

(2) 由于无须为其它系统开发各个数据交换程序, 成本低. 由于实用标准的数据交换程序极有可能已为每个系统所具有, 所以采用它就不需任何附加花费.

(3) 实用标准有最通用的软件包所支持, 支持该标准的市场很成熟.

由于受到初期开发者的经验及应用需求的限制, 实用标准不可能涵盖目前系统和应用程序的所有方面.

例如, AutoDesk 的 DBF, 在 GIS 间已成为广泛应用的数据交换格式, 但是它不能充分处理与空间实体相连的属性信息, 总会在交换过程中导致混乱和数据丢失。

2.1.3 基于国家的空间数据存储标准 为了更方便地进行空间数据交换, 减少空间数据交换损失的信息, 使之更加科学化和标准化, 许多国家和国际组织制定了空间数据交换标准, 如美国的 SDTF, 我国的 CNSDTF(“中华人民共和国国家标准地球空间数据交换格式”)^[4]。有了空间数据交换的标准格式后, 每个系统都提供读写这一标准格式空间数据的程序, 可以避免大量编程工作, 而且数据转换仅需要两次即可。从系统 A 的内部格式到标准的外部交换格式, 再从标准的外部交换格式到系统 B 的内部文件仅需两次转换, 省去为每种 GIS 软件都编写一个数据交换程序的步骤。其逻辑结构与系统间的数据流通与实用标准方法有些类似(如图 3)。避免了因缺少统一的数据标准而导致数据在转换过程中的翻来覆去。

2.2 数据互操作模式

所谓互操作, 是指不同的用户、应用程序及计算机系统能共享信息并一起工作, 而不管它们使用的桌面设备、联网硬件、各种协议以及运行数据的程序有什么差别^[5]。在不同系统之间, 利用通用数据的转换标准, 实现数据的间接转换。

数据互操作模式是 OpenGIS Consortium (OGC) 制定的规范。OGC 是为了发展开放式地理数据系统、研究地学空间信息标准化以及处理方法的一个非盈利性组织。GIS 互操作是指在异构数据库和分布计算的情况下, GIS 用户在相互理解的基础上, 能透明地获取所需的信息^[6]。

OGC 为数据互操作制定了统一的规范, 使得一个系统同时支持不同的空间数据格式成为可能。根据 OGC 颁布的规范, 可以把提供数据源的软件称为数据服务器(Data Servers), 把使用数据的软件称为数据客户(Data Clients), 数据客户使用某种数据的过程就是发出数据请求, 由数据服务器提供服务的过程, 其最终目的是使数据客户能够读取任意数据服务器提供的空间数据。OGC 规范基于 OMG(Object Management) 联盟发展的标准分布式对象结构 CORBA(Common Object Request Broker Architecture)、Microsoft 的 OLE/ COM 以及 SQL 等, 为实现不同平台间服务器和客户端之间数据请求和服务提供了统一的协议。OGC 规范正得到 OMG 和 ISO 的承认, 从而逐渐成为一种国际标准, 将被越来越多的 GIS 软件以及研究者所接受和采纳。目前, 还没有商业化 GIS 软件完全支持这一规范^[1]。

数据互操为多源数据集成提供了崭新的思路和规范, 将 GIS 带入了开放的时代, 为空间数据集中式管理、分布式存储与共享提供了操作的依据。OGC 标准将计算机软件领域的非空间数据处理标准成功地应用到空间数据上, 但是它更多地采用了 OpenGIS 协议的空间数据服务软件和空间数据客户软件, 对于已经存在的大量非 OpenGIS 标准的空间数据格式的处理办法还缺乏标准的规范。从目前来看, 非 OpenGIS 标准的空间数据格式仍然占据已有数据的主体, 而且非 OpenGIS 标准的 GIS 软件仍在产生大量非 OpenGIS 标准的空间数据, 如何继续使用这些 GIS 软件和共享这些空间数据, 成为 OpenGIS 标准面临的问题。

数据互操作规范为多源数据集成带来了新的模式, 但这一模式在应用中存在一定局限性^[7]: 首先, 为真正实现各种格式数据之间的互操作, 需要每种格式的宿主软件都按照统一的规范实现数据访问接口, 在一定时期内还不现实; 其次, 一个软件访问其他软件的数据格式时是通过数据服务器实现的, 这个数据服务器实际上就是被访问数据格式的宿主软件, 用户必须同时拥有这两个 GIS 软件, 并且同时运行, 才能完成数据互操作过程; 最后, 即使以后新建的 GIS 软件都支持 OpenGIS, 现有的 GIS 软件生产出来的空间数据也要转化到 OpenGIS 标准。

现在, 全世界很多家 GIS 软件公司正在制定开放性地理数据互操作协议(OGIS) 空间数据共享方案。主要目的是制定出一套各方能接受的空间数据操纵函数 API。遵循这一标准, 各厂商提供与这一 API 函数一致的驱动软件, 不需借助外部数据文件, 不同的软件就可以互相操纵对方的数据。

2.3 直接数据访问模式

数据直接访问方案这种刚萌芽的解决方案是指在一个 GIS 软件中实现对其他软件数据格式的直接访

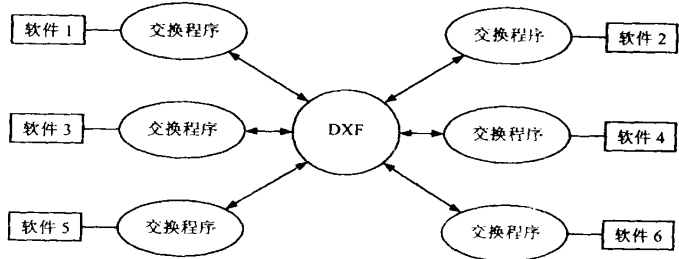


图2 基于数据交换程度的数据交换格式

问, 用户可以使用单个 GIS 软件访问、存取多种数据格式。目前仅有少数 GIS 软件开始了这种尝试, 如 Intergraph 的 GeoMedia 系列软件实现了对大多数 GIS/CAD 软件数据格式的直接访问, 包括 MGE、Arc/Info、Frame、Oracle Spatial、SQL Server、Access MDB 等。这种方案无须进行数据转换, 实现数据共享, 但它开发难度较大。

另一种途径是采用 Client/Server 体系结构, 将一个部门的所有空间数据和应用软件模块都共享一个平台。所有的数据都存在 Server 上, 各个应用软件都是一个 Client 端的程序, 通过这一平台向 Server 中存取数据。这是一种最好的空间数据共享方式, 任何一个应用程序所作的数据库更新都及时地反映在数据库中, 避免了数据的不一致问题。但是目前实现起来比较困难。因 GIS 软件厂商一般不愿意丢掉自己的底层, 而采用一个公共的平台。

3 结语

随着 GIS 技术的发展和应用范围的扩大, GIS 数据共享越来越引起人们的重视, 当前的数据共享方式弊端越来越明显: 数据转换模式虽然从某种程度上解决一些问题, 但是容易造成转换过程中的信息丢失, 而且违背了数据的分布和独立性原则^[5], 没有实现真正意义上的共享; OpenGIS 的本质还是各家 GIS 仍然用各自的格式, 另外, 由于 OpenGIS 函数集必须顾及各种 GIS, 所以, 空间数据互操作函数提供的信息和规模是最小的, 仍然存在着信息丢失的问题; 数据直接访问方式与 OpenGIS 相比, 是同一本质上的另外一种表现形式, 并没有从根本上解决问题。

最根本的途径是制定一个全球统一的数据格式, 通过将这一格式作为转换标准, 使之成为 GIS 数据交换的主流。各国政府加以宏观政策的引导和调控, 逐渐地, 越来越多的 GIS 公司就会将这一格式直接作为其核心数据格式。进而逐渐实现 GIS 数据格式的全球统一。实现这一目标, 统一标准要能够尽量概括去全空间对象的描述方法, 制定出全面的数据结构, 为不同领域和不同应用层次提供有效的、方便的统一标准。

除此之外, 应该制定和完善地理信息产业政策和法规, 以改变各部门之间数据共享代价过高的问题。这一切, 需要 GIS 工作者和相关组织的共同努力, 才能实现 GIS 数据真正意义上的共享。实现 GIS 数据共享, 收益的不仅仅是 GIS 软件开发商, 广大的 GIS 用户也将获益匪浅, GIS 的发展也将获得质的飞跃。

参考文献:

- [1] 宋关福, 钟耳顺, 刘纪远, 等. 多源空间数据无缝集成研究[J]. 地理科学进展, 2000, 2(19): 110-115.
- [2] 王磊, 周云轩. 21 世纪 GIS 发展趋势及误区分析[J]. 计算机工程与应用, 2002, 14: 54-57, 81.
- [3] 李红, 张献洲, 李永树. 常用 GIS 数据转换方式的比较研究[J]. 铁路航测, 2003, (1): 8-11.
- [4] 王艳东, 龚健雅, 黄俊韬, 等. 基于中国地球空间数据交换格式的数据转换方法[J]. 测绘学报, 2000, 29(2): 142-149.
- [5] 黄裕霞, 陈常松, 何建邦. GIS 互操作及其体系结构[J]. 地理研究, 2000, 19(1): 151-157.
- [6] 包世泰, 余应刚. 地理数据共享与互操作技术[J]. 测绘工程, 2000, 9(4): 32-36.
- [7] 梅士员, 江南. GIS 数据共享技术[J]. 遥感信息, 2002, (4): 46-51.

RESEARCH ON THE METHOD OF GIS DATA SHARING

CAI Aimin¹, ZHA Liangsong²

(1. Department of Geography, Chuzhou University, Chuzhou 239012, China; 2. College of Territorial Resources and Tourism, Anhui Normal University, Wuhu 241000, China)

Abstract: It's an emergency demand to realize the GIS data sharing. The reasons of the needs of data sharing are analyzed in this paper. Several methods are commented which are employed wildly nowadays. To realize the GIS data sharing, the fundamental method is to build a uniform data format in the world. Further more, some macro regulations and inductions should be given by the governments.

Key words: GIS; data sharing; data format