

关于我国地学数据整合与集成问题的思考

刘万崧¹ 丁 雷²

(吉林大学地球探测科学与技术学院, 副教授¹ 长春 130026; 吉林省勘查地球物理研究院, 高级工程师² 长春 130012)

〔摘要〕 基于对我国地学数据现状的分析, 阐述了地学数据整合与集成应借鉴的原则和地学数据库的建库内容, 对我国地学数据库建立与功能提升提出了建议。

〔关键词〕 地学原始数据 数据整合 地学数据库

〔中图分类号〕 K9 P22

〔文献标识码〕 A

〔文章编号〕 1000-7857(2004)12-0017-02

ON THE INTEGRATION OF GEOSCIENCE DATA IN CHINA

LIU Wan-song¹ DING Lei²

(1. School of Geo-Exploration of Science and Technology, Jilin University, Changchun 130026, China;

2. Jilin Geophysical Exploration Institute, Changchun 130012, China)

Abstract: Based on the results obtained from an analysis on the present situation of geoscience data in China, we propose the principle rulers for geoscience data integration and the contents of geoscience data base. Also we give suggestions on the establishment of geoscience data base and on the functions updating of the data base.

Key Words: original geoscience data, data integration, geoscience data base

CLC Numbers: K9 P22

Document Code: A

Article ID: 1000-7857(2004)12-0017-02

1 问题的提出

地学数据具有采集手段众多、数据类型复杂、数据所描述的地学现象时间和空间跨度大等特点。对于我国来讲, 国家曾花费大量资金用于地学数据的采集工作。但许多地学研究领域的原始数据资料是以图纸、模拟磁带、老式磁盘或光点相纸保存的, 有些资料保存时间已长达几十年, 长此下去会发生老化、去磁、沾连等问题而造成珍贵资料的丢失。另外, 我国从事地学调查研究的部门很多, 诸如地矿部门、石油部门、冶金部门、化工部门、核工业部门、水电建设部门、地震部门都曾进行过大量的地学调查工作, 这些资料保存在各有关单位或专家个人手中, 没有实现共享。随着单位隶属关系的变更、人员的流动, 也存在资料维护不当而损失、散失的危险。同时, 因原有数据格式以及存储方式不同, 对于大量的数据不能提供一个统一的数据接口、不能采用一种通用的标准和规范共享通用的数据源, 不利于研究人员使用, 资料价值亦得不到充分发挥。

目前, 国际上一些发达国家均采用数字技术建立了相应的地学数据库, 并对外开放, 通过互联网实现了国际间资料的交换与共享。近年来, 随着国内数字化技术和计算机网络技术的发展以及 GIS 技术的深入应用, 将现有地学数据和进一步地学勘探所获取的数据与国际通用数据格式接轨, 统一整合、集成在基于地理信息系统管理的数据库中, 是必要和可行的。这对于数据及时采集、整理和挖掘, 上网

发布、实现数据共享, 提高我国地学数据利用水平, 促进地学研究的发展有着重要意义。这些工作虽然在我国的某些部门已经启动, 但离实用目标尚有距离。

就当今数据整合与集成的理论、方法和技术水平而言, 还存在着一些问题。例如, 整合与集成的思想和方法多种多样, 但这些方法基本上都是源于某一个应用的需要而产生的, 由于这种多样性的产生, 使得人们对数据整合与集成的理解也是多种多样的, 有的理解甚至还是错误的; 对于行业来讲, 也尚未形成统一的框架体系。因此, 加强理论与方法、技术的研究, 制定行业的数据整合与集成的框架体系, 是非常重要的。

2 数据整合与集成的原则

GIS 应对用户提供集成的、统一的、安全的、快捷的信息查询、数据挖掘和决策支持服务。为了满足这个需求条件, 整合、集成后的地学数据必须保证一定的集成性、完整性、一致性和访问安全性。在这方面, 许多领域都提出了整合与集成的原则, 如集成性原则、完整性原则、一致性原则、访问安全性原则等(ICXO.COM), 地学领域可以借鉴。

(1) 集成性

数据经过整合、集成后, 应该达到查询一个综合信息不必再到各个业务系统进行分别查询和人工处理的要求, 只要在整合、集成后的数据信息仓库中就可以直接访问到, 即整合、集成后的综合信息仓库的数据是各数据的有机集成

收稿日期: 2004-09-09。

作者简介: 刘万崧, 男, 1959~, 吉林大学地球探测科学与技术学院(130026), 主要从事地理信息系统、地球物理的教学与科研工作, E-mail: liuwansong@email.jlu.edu.cn。

和关联存储(整合、发掘出各业务数据间的内在关联关系),而不是简单、孤立地堆放在一个数据库系统里。

(2)完整性

包括数据完整性和约束完整性两方面。数据完整性是指完整提取数据本身,约束完整性是指数据与数据之间的关联关系,是唯一表征数据间逻辑的特征。保证约束的完整性是良好的数据发布和交换的前提,可以方便数据处理过程,提高效率。

(3)一致性

不同信息资源之间存在着语义上的区别。这些语义上的不同会引起各种不完整甚至错误信息的产生,从简单的名字语义冲突(不同的名字代表相同的概念),到复杂的结构语义冲突(不同的模型表达同样的信息)。语义冲突会带来数据集成结果的冗余,干扰数据处理、发布和交换。整合、集成后的数据应该根据一定的数据转换模式和规则进行统一数据结构和字段语义编码转换。

(4)访问安全性

由于数据库资源可能归属不同的单位,各业务数据库系统有着各自的用户权限管理模式,访问和安全管理很不方便,不能集中、统一管理,所以为保证在访问异构数据源数据基础上保障原有数据库的权限不被侵犯,实现对原有数据源访问权限的隔离和控制,就需要设计基于整合、集成后的综合信息仓库的统一的用户安全管理模式。

3 地学数据库整合、集成、建库的内容

地学数据库的整合、集成和统一建库是一项涉及面广、任务量巨大的工作。目前,国土资源部正在进行中国岩石圈三维数据库的建设工作,正在研究和制定数据入库的技术标准和规范;科技产业部也开展了推动地学数据共享的数据库建库工作;此外还有一些部门也在开展类似的工作。这些工作在广度和深度上都还有很大的拓展空间。国家相关部门应做好协调工作,尽快研究制定适合我国国情又能较好地与国际接轨的相应的技术标准和规范。各相关部门应在此基础上进一步加强研究,不断扩充地学数据库建库的内容。

地学领域的研究工作涉及到对地质、地球物理、地球化学、遥感等多源信息的解释与利用。前提条件是对这些在格式和存储方式上的海量数据进行系统的标准化处理,按照统一的标准和规范进行整合与集成,存入基于GIS管理的数据库中。这些集成工作应包括以下几方面。

(1)空间特征数据与属性特征数据的集成

目前人们已经尝试将空间特征数据与属性特征数据都存放到关系型数据库中。面向对象技术的出现,使得用面向对象的数据库来统一管理上述两种数据成为发展的趋势。

(2)多源空间数据集成

地学空间数据量大、数据来源广泛、数据类型复杂,对这些数据尽可能准确、及时地采集、整理和挖掘,是当今GIS界的研究热点和技术难点。

(3)基于空间元数据的集成

地学空间数据集成是以分布式网络数据为处理对象,以应用项目为集成目标的动态数据处理过程,各个层次的元数据通过集成系统各部分运行状态和处理对象的特征信息来控制集成数据的功能流向,以实现数据集成的最终目标。

(4)与应用模型的集成

基于GIS管理的地学空间数据库为利用GIS进行空间分析提供数据支持。应用分析模型是解决特定的分析预测问题的数学计算模型,将GIS与应用模型予以集成可以提高GIS的分析功能,更最大限度地发挥地学数据库中数据的作用。

(5)与知识规则库集成

从GIS数据库中挖掘知识是GIS重要的应用方向之一。应用认知与知识发现技术从GIS数据库挖掘知识的内容包括数据精练、数据更新、自动建立空间知识库,并可与专家系统技术结合,使GIS发展成为智能化的空间信息系统。

4 地学数据库建立与功能提升的几点建议

前面提到的地学数据整合、集成原则以及集成工作,是当今GIS研究领域的重要研究内容。结合我国地学数据的实际情况以及目前开展的部分地学数据库建库工作,提出如下建议。

(1)由国家有关部门牵头,开展对各行各业现有地学数据库情况的调查。制定相关的政策,调动各行业参加地学数据库建设的热情。

(2)尽快制定国家级地学数据库总库与子库框架体系,以及相应数据入库的技术标准和规范。

(3)国家级的地学数据库应包括地质、地球物理、地球化学、遥感等原始资料组成的原始数据库。多年来各部门都进行了一定的地学数据研究处理工作,积累了一批成果,这些成果经专家筛选后可以建立成果库,以及描述原始数据库和成果库的元数据库。

(4)按照相应标准进行资料数字化工作和数据转换的工作,特别是加强数据转换方面的研究。如吉林大学有关人员进行了中国地学断面图形数据库的研制工作,在工作中较好地实现了MAPGIS系统平台上图形数据库文件向ArcInfo系统平台上图形数据库文件的转换。解决了常规转换中信息丢失的问题。

(5)在进行新数据采集工作时,严格按照相关的数据格式要求记录,并将数据及时提交入库。

(6)规范化的地学数据库能够对不同的地理信息系统提供有利的数据支持和存储其成果。与应用模型和知识规则库的结合是地学数据库功能提升的内容。例如,调用数据库中的地质、地球物理、地球化学、遥感资料,在成矿地质先验前提下,经综合矿产预测系统的标准化处理,以权威领域专家的知识经验有效地提取出特征信息,并对多学科信息进行自动分析与综合,从而预测出新的找矿靶区;处理成果亦可以存入成果库。

总之,建立国家级的地学数据库,完成对多源地学数据的整合与集成,是建立数据共享开放环境的需要。虽然这是一项复杂的工作,但对于提高地学研究工作效率与水平有着重要的意义。

参考文献(References)

- [1] 阎国年. 地理信息系统集成原理与方法 [M]. 北京: 科学出版社, 2003
- [2] 龚建雅. 当代GIS的若干理论与技术 [M]. 武汉: 武汉测绘科技大学出版社, 1999

(责任编辑 肖庆山)