

文章编号: 1007-6301 (2002) 04-0365-09

# 基于元数据和数据集管理的 应用模型研究

赖志斌, 夏曙东, 王 浒, 承继成  
(北京大学遥感与地理信息系统研究所, 北京 100871)

摘要: 元数据是解决数据共享的技术手段之一, 现有以元数据为基础的管理模型和管理系统主要针对元数据本身的管理, 元数据和数据集在管理上是相对独立的, 这直接导致了元数据和数据集的不一致性。在元数据与数据集内在联系的基础上, 本文提出了实现元数据和数据集一体化管理的应用模型, 该模型由综合管理层、服务层、应用层三层体系结构组成, 通过服务流、数据流、元数据流的交互, 实现元数据和数据集的同步一体化管理。从数据集对象的建立、数据集对象的查询和获取、数据集对象的变更三个应用方面, 文章详细阐述了该应用模型的实现流程。最后, 采用基于元数据和数据集一体化模型开发的 I-IMS 系统, 在厦门市环保局元数据管理系统的建设中进行了典型应用研究。

关 键 词: 元数据; 数据集; IIMS

中图分类号: N941 文献标识码: A

## 1 引言

元数据是数据的数据, 它描述了数据的内容、质量、情况和其它特性。元数据帮助人们定位和理解数据<sup>[1,2]</sup>。数据集是元数据所描述对象数据的最小单元, 它可以是单个文件、多个文件或者数据库数据。元数据根据它的空间性可以分成空间元数据和非空间元数据, 根据各个行业的业务数据类型不同, 形成各种行业元数据, 如在环保行业形成的是环保元数据、在国土行业形成的是国土元数据等。元数据技术是数字地球<sup>[3~5]</sup>中的一项关键技术, 是实现信息资源共享集成的基础性支持手段<sup>[6,7]</sup>。

进行元数据研究的一项基础性工作是如何建立起元数据标准体系<sup>[8]</sup>, 除了对元数据标准体系的研究之外, 许多学者对建立在元数据基础上的数据共享体系<sup>[9]</sup>以及分布式数据的管理模型展开了深入的研究, Douglas M. Flewelling<sup>[10]</sup>提出了文件档案管理支持下的数据集获取模型, 李浩川<sup>[11]</sup>提出了元数据的组织方法及数据的管理方法, 陈爱军、黄晓斌<sup>[12]</sup>提出了元数据管理的概念模型。在元数据管理的概念模型研究的基础上, 沈体雁、程承旗<sup>[13]</sup>

收稿日期: 2001-08; 修订日期: 2002-04

基金项目: 国家“九五”重大科技攻关项目子专题“超媒体空间信息系统及技术集成研究”(97-79-04)

作者简介: 赖志斌(1973-), 北京大学遥感与地理信息系统研究所博士研究生, 主要从事网络 GIS、国家信息基础设施和元数据应用方面的研究, 已发表论文 10 余篇。

E-mail: zhibinlai@263.net

提出了基于空间元数据的网络数据管理模型,李琦、陈爱军、夏曙东<sup>[14]</sup>提出了空间元数据网络管理系统的体系结构。以上的这些模型或系统构架主要是在 web 浏览器的支持下,用户对数据集进行元数据的注册,注册结果提交到元数据库中,并通过 ftp 或者 HTTP 的方式进行数据集地址的描述,反映数据集的提供者或发布地址。用户通过元数据管理系统进行数据集的查询,通过元数据描述项中的 FTP 或者 HTTP 地址获取数据集。

在以上的元数据管理模型中,元数据的管理和数据集的管理是独立的,虽然数据集和元数据存在一一对应关系,但是两者的管理本质上却属于互不相关的独立系统,缺乏直接联系。元数据和数据集的关联只是通过元数据中位置信息的描述对数据集的位置进行解释,这直接影响到元数据的应用效果。它的不足之处体现在两个方面:一方面是元数据与数据集之间缺乏同步性,由于元数据管理和数据集管理属于两种不同的管理体系,当数据集内容发生变化时,元数据内容缺乏相应的响应和更新机制,这直接影响了元数据内容的准确性,即元数据的数据质量;另一方面,数据集的直接获取困难,例如 FGDC 的数字地理空间元数据标准,对数据集的位置信息的定义方法是通过发布信息项中的发布人地址描述来实现的,但当数据集的内容发生变化时,用户获取的是与元数据内容不相符合的数据集,如果数据集管理系统由于故障问题而停止了信息提供服务,在元数据管理端无法响应该类情况,用户即使获取了元数据内容也无法获取到相应的数据集。

元数据的其中一项很重要的应用服务内容是提供数据的查询支持,用户查询元数据的最终目的希望获取该元数据所对应的数据集内容,因此,如何实现元数据与数据集的一体化管理,保持元数据和数据集内容的同步和一致,实现元数据和数据集从建立、查询到获取过程的协同功能操作,是面向信息资源共享所需解决的一项关键技术问题,也是元数据从单纯的目录内容管理服务走向深层次实用化的关键技术途径。

## 2 元数据和数据集一体化应用模型

元数据是在对数据集进行描述的基础上产生的,因此元数据和数据集存在一一对应的关系,这就为建立元数据与数据集一体化管理模型的建立提供了可行性前提。基于元数据和数据集在管理上相对独立的现实情况,一种有效的途径是在元数据管理和数据集管理的层面上提供统一的综合管理层,以此来协调元数据服务和数据集服务的一致性和同步性。元数据和数据集自身的管理可以采用分布式的管理体系,对用户而言,只需要保持他们对元数据服务器和数据集服务器实时访问的畅通途径。在这种思路下,我们提出了如图 1 所示的元数据和数据集一体化管理的应用模型。

该模型可以划分为 3 个层面,即:综合管理层、服务层和应用层。这 3 个层面交互的信息流包括:服务流、元数据流、数据流。

综合管理层分布的是核心服务器,它主要是对元数据服务器和数据集服务器的注册管理,保持元数据服务器与数据集服务器的存在关系以及用户权限的安全控制,进而保证在元数据服务器中的元数据内容 and 数据集服务器中的数据集内容的同步和一致。它与服务层通过服务流进行交互。

服务层包括两种类型的服务器,一类是元数据服务器,另一类是数据集服务器。元数据服务器管理的是元数据库,数据集服务器管理数据集内容,由于一类元数据标准所对应

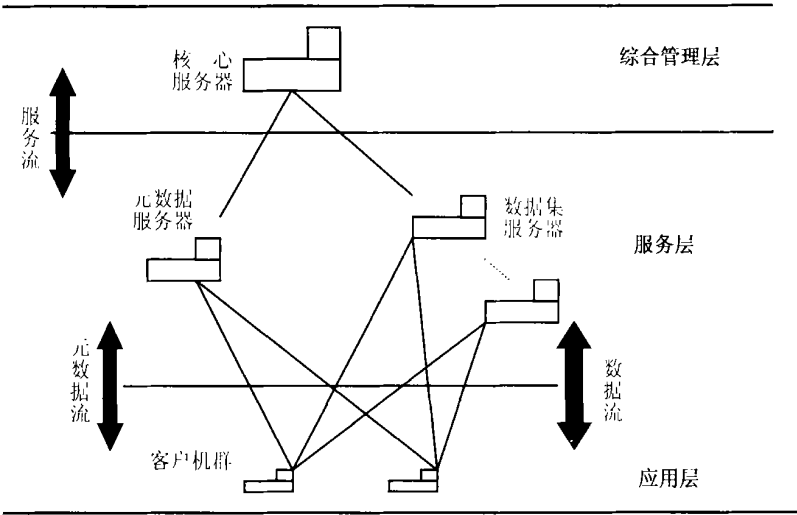


图 1 元数据和数据集一体化管理模型体系结构图

Fig. 1 Application model structure based on Metadata and Dataset

的数据集的内容具有多样性和数据量大的特点，数据集的管理采用分布式的管理模式，元数据服务器和数据集服务器可以是一对多的关系。

应用层主要是客户机群，客户机通过客户端系统对服务层的元数据服务器和数据集服务器进行访问。访问元数据服务器主要实现元数据内容注册、提交，对元数据内容的匹配查询、查询结果的显示等功能；访问数据集服务器主要实现数据集的提交、预览、下载等功能。应用层的客户机与服务层中的元数据服务器通过元数据流进行交互，与服务层中的数据集服务器通过数据流进行交互。

下面从数据集对象的建立、查询和获取、更新三个方面进一步阐述该模型的应用流程。

2.1 数据集对象的建立

数据集对象的建立指的是应用层的客户机在确定数据集对象的情况下，创建该数据集对象的元数据内容，并将数据集对象提交至数据集服务器和将对应的元数据内容提交到元数据服务器中的过程。该功能操作也可称为数据集与元数据的注册，其实现流程如图 2。

应用层的客户在建立数据集后，通过客户端应用程序，在综合管理层的支持下，首先进行元数据服务器的定位、数据集服务器的定位以及用户的合法性检查等三项功能操作。

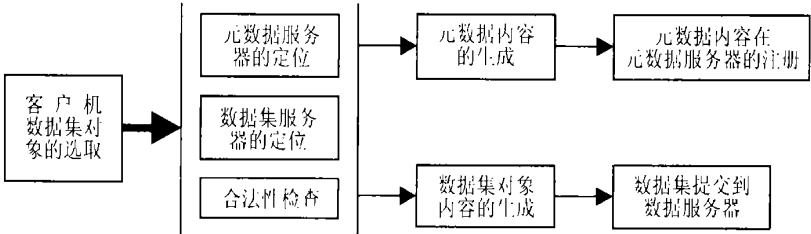


图 2 数据集对象建立流程图

Fig. 2 Creation flow of dataset object

元数据内容的生成是指对元数据标准体系中具体内容项的填写,元数据内容在元数据服务器中的注册指的是元数据内容在元数据库中的存储。客户进行这两项元数据的功能操作时,与服务层中的元数据服务器进行的是元数据流的交互。

元数据内容的生成建立在数据集对象内容生成的基础之上,数据集提交到数据服务器是指数据集对象在数据服务器中的定位存储过程,用户对数据集进行功能操作时,与服务层的数据集服务器进行数据流的交互。

从以上分析可以看出,在数据集对象的建立全过程,元数据和数据集的功能操作是同步进行的。

## 2.2 数据集对象的查询和获取

数据集的查询和获取内容包括:在元数据服务器的支持下,客户机通过元数据搜索引擎查询到满足条件需求的元数据;通过综合管理层的支持,结合元数据内容中对数据集对象的存储位置的描述,客户端系统定位存储该数据集对象的数据集服务器,并从数据集服务器上获取到该数据集对象。它的实现流程如图3。

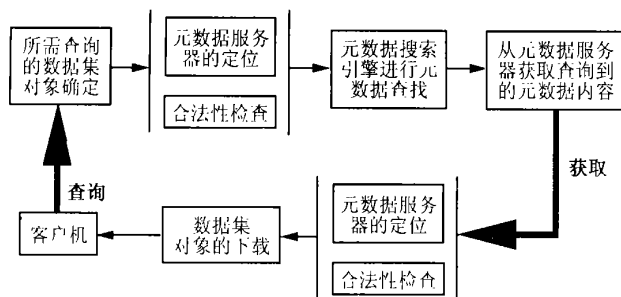


图3 数据集对象查询和获取流程

Fig. 3 Querying and getting flow of dataset object

应用层的客户端进行数据集查询和获取的功能操作流程包括元数据操作流程和数据集操作流程两个部分内容。

元数据操作流程是:通过综合管理层的支持,首先实现元数据服务器定位以及用户的合法性检查,接着应用层客户机中的元数据搜索引擎进行元数据内容的匹配查询,并获取到该查询结果,客户可以查看元数据的具体内容项。在整个流程中,用户层的客户机与服务层中的元数据服务器进行的是元数据流的交互。

数据集操作流程是:通过元数据结果内容项中对数据集的具体位置的描述信息,在综合管理层的支持下,首先实现数据集服务器的定位和用户权限的合法性检查,用户从数据集服务器中下载该数据集对象,实现对数据集对象的预览、显示以及客户端的其他功能操作。在该流程中,用户层的客户机与服务层中的数据集服务器进行的是数据流的交互。

从以上流程分析可以看出,在综合管理层的支持下,数据集对象的查询和获取的整个过程始终保持着对元数据操作流程和对数据集操作流程的连贯性,进而保证了元数据和数据集内容的一致性。

## 2.3 数据集对象的变更

数据集对象的变更操作是元数据管理系统的一项关键技术环节,也是一个技术难点问

题，它的功能操作的完善与否，将直接导致元数据和数据集对象能否保持一致。在本应用模型中，数据集对象的变更包括数据集对象在数据集服务器上的更新和与之相对应的元数据内容在元数据服务器中的更新。

数据集对象的更新首先要进行数据集对象的获取操作，用户对数据集对象内容完成变更操作后，将进行元数据内容变更及注册以及数据集对象的变更提交功能流程。在这个功能操作流程中，关键点是数据集对象存储方式，如果数据集对象在数据集服务器上采用“覆盖型”的存储方式，则元数据的操作只是对历史元数据的修改覆盖。如果数据集对象在数据集服务器上采用“增加型”的存储方式，则元数据的操作是增加新的元数据内容，保持历史元数据内容不变。它的实现流程如图 4 所示。

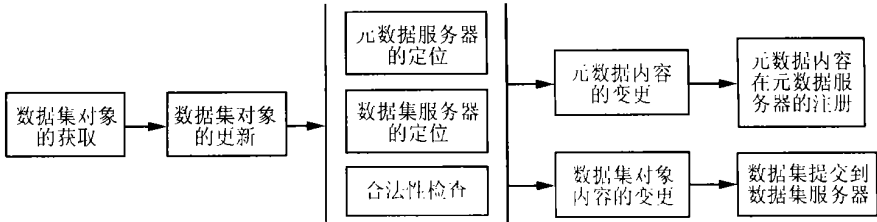


图 4 数据集对象的变更流程

Fig.4 Revising flow of dataset object

应用层的客户在客户端完成了对数据集对象更新操作后，将进入数据集对象的变更流程，该流程包括元数据操作流程和数据集操作流程两个部分的内容。这两个流程的操作与数据集对象的建立流程相类似。所不同的是在综合管理层支持下，首先判断数据集在数集服务器中的存储方式，该存储方式是由客户端进行定义的。当用户端选择“增加型”的存储方式时，整个操作流程与数据集对象的建立流程完全相同；当用户端选择“覆盖型”的存储方式时，数据集提交到数据集服务器是取代了原有的数据集对象，历史数据集对象将不保留，同时，元数据内容在元数据服务器上的注册也是取代了原有的元数据记录，历史元数据记录将不保留。

从以上流程分析可以看出，在此应用模型的流程操作体系下，元数据和数据集的内容始终保持着同步和一致，尤其涉及到数据集内容和元数据内容的调整时，在综合管理层的支持下，两者的同步操作避免的元数据或者数据集内容的单方面变更。

### 3 应用研究

厦门市环保局是国内最早开展环保信息化工作建设的单位之一。在多年的信息化建设过程中，该局积累的大量的环保数据，这些数据资源在环保局信息中心进行集中管理。数据资源主要分成三大类内容：空间数据、办公业务数据以及环境监测数据。空间数据包括整个厦门市的 1 2000、1 500 的基础地形图库、厦门市历年的 TM、SPOT 卫星影像图库、2001 年的厦门市单波段 1 米分辨率和红、绿、蓝、近红外四个波段的 4 m 分辨率的 IKNOS 卫星影像图库<sup>[16]</sup>。办公业务数据包括整个厦门市环境保护局办公自动化系统中运行的数据

内容。环境监测数据包括历年厦门市环保局的监测数据，如：不同时间段的厦门市重点污染源监测数据、大气监测数据、水污染监测数据等等。目前整个厦门市环保局信息中心所管理的这三大类数据的容量近 200G，由于缺乏有效的数据共享的管理体系，数据资源的检索、查询和环保局各个处室的共享应用非常不方便。因此，建立起一整套行之有效的元数据管理系统，为厦门市环境保护局的各个业务处室提供数据资源的共享应用体系，是提出和建立厦门市环保局元数据管理系统的出发点和直接目标。

厦门市环保局元数据管理系统是基于元数据体系的分布式数据管理系统。该系统采用信息综合管理系统 IIMS 作为支持平台。IIMS 是基于元数据和数据集一体化管理模型的分步式信息资源管理系统，它采用 JAVA / CORBA<sup>[9]</sup> 进行开发，直接支持 TCP/IP 进行网络连接，脱离了浏览器而采用 IIMS 前台客户端系统的操作方式。该系统由服务器管理，元数据管理，数据服务器管理和应用管理四大部分构成。图 5 是 IIMS 系统中的综合管理层系统界面图。

厦门市环境保护局元数据管理系统的总体框架采用元数据与数据集一体化模型的三层体系结构：在综合管理层是厦门市环保局的信息综合管理服务器；在服务层包括元数据服



图 5 IIMS 综合管理层系统界面

Fig. 5 Interface of integrating management level of IIMS

务器、空间数据集服务器、办公数据集服务器和监测数据集服务器, 元数据服务器管理整个环境资源元数据标准和元数据内容, 空间数据管理厦门市的基础地形图库和影像库数据内容, 办公数据集服务器管理厦门市环保局的办公自动化的数据和日常的文档数据, 监测数据集服务器管理厦门市环保局的环境监测数据; 在应用层, 各办公部门安装 IIMS 客户应用系统。整个体系构架如图 6。

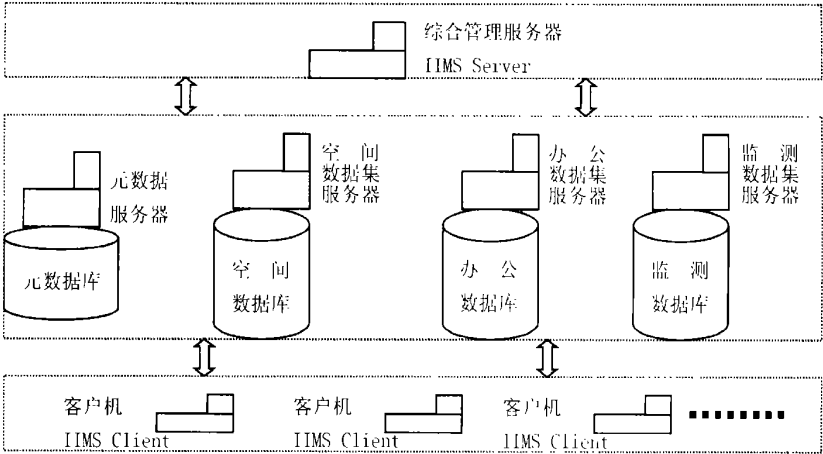


图 6 元数据管理系统三层体系结构图

Fig. 6 Three level structure of metadata M anagement system

厦门市环保局的元数据管理系统中元数据标准包括空间元数据标准、办公元数据标准、监测元数据标准。目前, 第一阶段实现了空间元数据标准的建设工作。空间元数据标准选取的是美国 FGDC 的数字地理空间元数据标准, 考虑到该标准的内容项数目教多和环保行业特殊性, 实际操作中是提取出针对环境资源特点的数据内容项形成环保空间元数据标准。

整个系统围绕元数据管理展开, 如在环境监测数据管理过程中, 信息中心的工作人员数字化一幅厦门市重点污染源监测图后, 启动 IIMS 应用系统进行该图的元数据注册工作, 并将该图提交至空间数据集服务器。其它处室的工作人员通过 IIMS 的元数据查询引擎查找到该图的元数据内容, 并直接从空间数据集服务器中获取该图, 系统根据数据的类型启动 GIS 系统 (ArcView) 显示该图内容, 在本地进行审查和修改操作, 完成后将图进行重新注册和提交的工作。信息中心的制图工作人员通过 IIMS 系统进行该图的查找和获取工作, 制图输出。在这些业务工作中, 工作人员通过该系统可以实现对厦门市的环境图库便捷的检索、获取、更新等功能操作, 元数据和数据集一体化的管理模式在厦门市环保局的元数据管理系统中得到成功的应用。

4 结论和展望

通过厦门市环保局的元数据管理系统的典型研究表明, 元数据和数据集一体化管理模型具有可行性。目前该模型只涉及到一类元数据服务器的类型, 由于元数据的标准是随行

业的不同具有多样性, 每类元数据标准的都对应分布式数据集对象, 数据集对象之间往往又存在相互交叉的情况, 因此, 如何实现元数据集服务器和数据集服务器同时在分布式环境下的一体化管理, 将是有待于进一步研究的内容。

## 参考文献:

- [1] ISO/TC211 Metadata Working Document: Geographic Information, 1997, <http://www.fgdc.gov/metadata/metadata.html>
- [2] ISO/TC211 Metadata Working Document: Geographic Information, 2000, <http://www.fgdc.gov/metadata/metadata-workbook.html>
- [3] AL Gore. The Digital Earth: Understanding our Planet in the 21st Century[M]. Los Angeles, California, January 31, 1998. <http://www.digitalearth.net>
- [4] 郭华东, 扬崇俊. 建设国家对地观测体系, 构建“数字地球”[J]. 遥感学报, 1999, 3(2): 90-93.
- [5] 赵永平等. 数字地球的体系研究[J]. 地理科学进展, 1999, 19(1): 38-45.
- [6] 马荣华, 黄杏元, 蒲英霞. 数字地球时代 3S 集成的发展[J]. 地理科学进展, 2001, 20(1): 89-95.
- [7] 齐清文, 裴新富. 多源信息的集成与融合及其在遥感制图中的优化利用[J]. 地理科学进展, 2001, 20(1): 36-43.
- [8] 李军, 周成虎. 地球空间数据元数据标准初探[J]. 地理科学进展, 1998, 17(4).
- [9] Harlan J Onsrud, Gerard Rushton eds. Metadata Issues in Sharing Geographic Data[A]. In: Sharing Geographic Information[C]. New Brunswick: Rutgers, The State University of New Jersey, 1995. 499.
- [10] Flewelling D M, Egenhofer M J. Using digital spatial archives effectively[J]. *INT. J. Geographical Information Science*, 1999, 13(1): 1-8.
- [11] 李浩川. 基于网络环境下的地理空间数据组织、管理及其在可持续发展决策支持系统中的应用[硕士学位论文][A]. 北京: 北京大学, 1998.
- [12] 陈爱军, 黄晓斌. 数字地球中的元数据管理模型研究[J]. 中国图形图像学报, 1999, 11(4): 996-1000.
- [13] 沈体雁, 程承旗. 基于空间元数据的分布式地理数据管理模型及应用研究[J]. 测绘通报, 1999, 7: 34-36.
- [14] 李琦, 陈爱军, 夏曙东. WebGIS 中地理空间 Metadata 管理系统的研究与设计[J]. 中国图象图形学报, 2000, 10(5): 811-816.
- [15] Joseph Weber. USING JAVA TM 1.1 (THIRD EDITION), 1996.

## Research of Application Model Based on Management of Metadata and Dataset

LAI Zhi-bin, XIA Shu-dong, WANG Hu, CHENG Ji-cheng

(RS and GIS Institute of PeKing University, Beijing 100871 China)

**Abstract:** Metadata is one of the valid technique for data sharing. Current Metadata management models and management systems focus on the Metadata, the management of Dataset is independent, and there is no direct connection between Metadata and Dataset, it leads to the inconsistency of Metadata and Dataset. Due to the direct relationship of Metadata and Dataset, the management model based on Metadata and Dataset is given, it consists of three level including integrating management level, service level, application level and solve the independent management of Metadata and Dataset. Integrating management



level exchanges services flow with service level. Service level manages the distributing Metadata servers and Dataset servers, and it exchanges Metadata information flow and Dataset information flow with application level. Application level consists of customer machines.

In addition, the application flow of this model is explained in three parts. The first part is the construction of Dataset object. it includes the filling and registering Metadata, submitting the Dataset to Dataset server. The second part is Dataset object querying and getting. it includes searching the Dataset based on the Metadata searching engine, and downloading the Dataset from Dataset server. The third part is Dataset object revising. it includes the Metadata revising in Metadata Server and Dataset revising in Dataset server consistently.

In the end, with IIMS which is based on the integrating application model, typical application research is done in Xiamen Environment Bureau's Metadata management system.

**Key words:** Metadata; Dataset; IIMS

\* \* \* \* \*

紧 急 征 订

《地理科学进展》是中国科学院地理科学与资源研究所主办的综合性学术期刊,为更及时地反映地理科学的研究成果,经国家科技部批准,于2002年起正式改为双月刊,每期96页,国内刊号:CN 11- 3858/P,国际刊号:ISSN 1007- 6301,国内邮发代号已改为2- 940,每期定价18元,全年共计108元,全国各地邮电局均可订阅。如有漏订者请速与科学出版社期刊部或《地理科学进展》编辑部联系。

联 系 人: 沈志华、何丽光  
地 址: 北京市东黄城根北街16号  
电 话: 010- 64034563; 64017153  
邮政编码: 100717  
开户单位: 科学出版社期刊部  
开户银行: 北京市商业银行沙滩支行  
帐 号: 03252001201057743- 15