

分类号\_\_\_\_\_

密级\_\_\_\_\_

UDC\_\_\_\_\_

编号\_\_\_\_\_

# 中国科学院研究生院 硕士学位论文

中国虚拟天文台的数据库和数据服务

桑 健

指导教师\_\_\_\_\_ 赵永恒 研究员 \_\_\_\_\_

\_\_\_\_\_ 中国科学院国家天文台 \_\_\_\_\_

申请学位级别 硕 士 学科专业名称 天文技术与方法

论文提交日期 2004年6月 论文答辩日期 2004年7月

培养单位\_\_\_\_\_ 中国科学院国家天文台 \_\_\_\_\_

学位授予单位\_\_\_\_\_ 中国科学院研究生院 \_\_\_\_\_

答辩委员会主席\_\_\_\_\_ 阎保平 \_\_\_\_\_

## 摘 要

随着科学技术的进步，特别是望远镜制造技术、大面积探测器阵列技术和一系列计算机科学技术（计算技术、存储技术和网络技术）的飞速发展，一大批巡天设备的投入运行，天文数据正以前所未有的速度增长，目前天文数据量已经达到 PB 量级，天文学开始进入一个信息丰富的新时代。

如何实现全球海量、高度分布、多波段天文数据的高度共享和融合，实现包括数据资源在内的各种资源的无缝统一访问，实现海量数据的高效处理、分析和挖掘，以发挥出天文数据库的全部科学价值，产生出新的重大科学产出，成为目前急需解决的问题。旨在解决的这些问题的虚拟天文台构想应运而生并迅速得到全世界天文界的关注和重视，各国的虚拟天文台项目纷纷产生，中国虚拟天文台也是其中一员。

作为中国虚拟天文台的首批技术实施者，我们在比较广的领域进行了研究探索：

首先是在天文数据库的构建方面进行了研究。天文数据库是虚拟天文台最重要的基础设施，但是以往的天文数据库往往缺乏对元数据的收集和管理，已经不能满足虚拟天文台对天文数据库的信息完整性和互操作性的需求。我们在适应中国虚拟天文台的天文数据库的结构方面做了一些积极研究探讨。我们利用 CDS 释放的星表数据作为数据源，设计和实现了中国虚拟天文台的天文星表数据库管理系统，系统采用 MySQL 数据库系统软件，目前入库的星表达达到 112 个，数据量达 311GB。天文星表数据库系统包括“数据”库和“元数据”库两部分，我们着重于对星表元数据库结构的设计，借鉴和采纳了虚拟天文台资源元数据描述标准和 VOTable 格式对数据字段的描述，以及 UCD1 和 UCD1+ 标准。保证了对星表元数据和数据字段元数据的描述的完整性，为搭建虚拟天文台的数据服务和实现自动交叉证认提供了坚实的基础。另外，为了提高数据入库的效率，我们制订了一套数据入库的规则，并用 Java 语言实现了遵照此规则的星表数据自动入库程序。

其次，我们利用基于开放网格服务架构（OGSA）的 Globus 工具包，并结

合现有的天文数据标准，搭建了多种数据访问服务。首先，我们设计了星表 Cone-Search 访问接口，并基于不同的星表数据库系统搭建了多个星表的 Cone-Search 服务，这是中国虚拟天文台实现的第一个网格服务，初步实现了不同本地环境下分布式天文星表数据的统一查询访问。证实了利用网格技术解决海量分布天文数据融合的可行性。其次，为了实现更加复杂的查询功能，和融合现有的关系型数据库中的数据资源，我们又设计了星表查询访问接口，引入了 ADQL 作为描述查询要求的统一语言，并初步搭建了基于 MySQL 数据库的星表查询访问服务。最后，我们在天文图象的统一访问上也做了积极的探讨，设计了一个简单的天文图象访问接口并实现了 DSS-I 天图访问服务。

还有，在 VOTable 格式数据的显示上，我们首先开发了用于将其转换为 CSV 和 HTML 的 XSLT 样式表，然后提出了在 Web 环境下将 VOTable 文档以 HTML 或图形显示的几种技术解决途径，并逐一实现。网格环境下的数据服务并不能被天文学家直接利用，必须为他们设计一个基于 Web 的查询可视化查询界面，通过这个界面，可以访问到分布各地的数据服务，真正实现了数据的统一访问。我们基于前面提到的星表 Cone-Search 服务实现了星表图形可视化查询界面。

通过在以上多个方面的研究和开发工作，我们为虚拟天文台的实现奠定了基础，证实了通过网格技术搭建虚拟天文台框架结构的可行性，并虚拟天文台的进一步发展开辟了一个广阔的前景。

**关键字：**虚拟天文台 数据库 数据访问服务 VOTable 可视化

## 第一章 虚拟天文台概述

### § 1.1 信息时代的天文学<sup>[1]</sup>

近十多年来,天文学正进行着革命性的变化,这就是天文学已经步入了信息数据丰富的新时代<sup>[2,3]</sup>。这一变化是由科学技术的飞速进步推动的,其中主要的是以下几个技术——望远镜设计制造技术、大面积探测器阵列制造技术和计算机科学技术(主要包括计算技术、存储技术和网络技术)<sup>[1,4]</sup>。

望远镜技术的进步使得人们可以建造大型的地面和空间的天文设备,这在伽玛射线、X射线、紫外、可见光和红外波段对宇宙进行观测开辟了广阔的前景,一系列各波段的空间天文台投入建设和运行;同时也推动了新一代大口径地面光学和射电望远镜的建造。望远镜制造技术事实上反映了当前时代工程科学所能达到的最高设计水平<sup>[5]</sup>。在光学波段,由于主动光学技术、自适应光学技术、薄镜面拼接技术以及其他先进技术的不断进步和采用,人们可以制造出口径越来越大,功能越来越好、观测效率越来越高、数据获取量越来越大的天文观测设备,目前口径最大的光学望远镜为10米的凯克望远镜,100米口径望远镜OWL的计划也已形成<sup>[6]</sup>,并有希望在不久的将来实现。在射电波段,目前最大的综合孔径望远镜为美国的甚大望远镜VLA,而平方公里阵(SKA<sup>[7]</sup>)射电望远镜也在积极筹划中。

大面积探测器技术也在突飞猛进的发展,在可见光波段和红外波段,现在望远镜探测器已经采用CCD替代了过去的照相底片,具有更大面积和像素数、更高量子效率、更宽的响应波段的特性的CCD已经出现和应用。目前拼接CCD像素数已经达到 $10^8$ 量级,而 $10^9$ 像素的拼接CCD也将很快出现。CCD的像素数目反映了采集到数据的大小,比如一个4K\*4K的CCD,每个像素读出值用2个字节表示,那么一次读出数据大小就是32MB。

更多的地面和空间大口径和配备有复杂仪器的天文观测设备的投入观测的必然结果是数据流的极大增加。而观测模式的转变也是造成数据量庞大的重要原因,尽管新的地面和空间天文台还将继续给单个观测者/单计划观测模式的研究一定量的观测时间,但更多的时间是进行大规模的,有计划的,经常是多波段的

巡天观测，涉及到大量的合作者。这些大的巡天计划产生出海量质量均匀、标准统一的天文数据，总数据量通常要用TB ( $10^{12}$ Byte) 来衡量。目前正准备放弃的哈勃空间望远镜近 14 年的获得了大约 8TB 的观测图像，而正在进行的斯隆数字巡天 (SDSS<sup>[8]</sup>) 5 年观测计划会产生大约 40TB 的原始观测数据，最终数据产品的数据量也会达到 15TB。中国正在建造的LAMOST<sup>[9]</sup>望远镜也是海量巡天数据的生产者，它每年产生的原始数据量在 7TB 以上，6 年观测将会产生超过 40TB 数据，观测 2 千万颗天体的光谱。而美国计划建造的“大口径概要巡天望远镜 (LSST<sup>[10]</sup>)”观测数据量将会达到每天 16TB 的量级，每两周对全天观测一遍，5 年计划观测总数据量将会达到 15PB ( $15 \times 10^{15}$ Byte)<sup>[11]</sup>。这种观测模式的变化，不仅由于新设备的威力的提高允许更快速地获得这些数据，还由于计算机的硬件和软件可以快速地进行数据采集、处理。而存储技术的进展使得存储海量数据变得经济可行。

飞速发展的互连网技术使得各国的天文学家可以方便获取这些数据，虽然现在通过一般的网络传输海量的天文数据还不太可行（传输速度太慢），但是未来的网络会快的多，这种数据传输速率加上地面和空间设备的高效率的数据采集，使得在不同地点的大量数据交换成为可能，这将使世界各地的天文学家都能够访问和使用这些天文数据，从而具有巨大的科学产出的潜在意义。

互联网时代的天文数据有着其他学科数据所无法比拟的特点<sup>[1,4]</sup>：

1. 天文数据绝大部分是开放数据。天文数据不具备商业价值，天文界一个良好的传统是将得到的观测数据一年后向全世界天文学家和公众开放，这样既保证了数据贡献者有一年的时间分析数据得到科学产出和回报，也使全世界天文学家有机会使用这些数据。天文学的数据开放和共享机制在各类学科中是独一无二的。
2. 天文观测波段已经从可见光波段拓展到全电磁波段，从射电波段、红外、可见光、紫外到 X 射线、 $\gamma$  射线，都能对天体进行探测。这就使得我们在每一个波段上都拥有丰富的观测数据。

天体辐射的能量是分布在非常广泛的波段上的，从射电到红外、光学到紫外、X 射线甚至到伽马射线，因此每个波段上的观测都带来

了有关天体本质的重要信息。而同样一个天体在不同波段上的表现是可以完全不同的<sup>[12]</sup>，如蟹状星云（图 1）的光学图象显示出了电离氢的分布，射电图象显示了中性氢的分布，红外图象显示了尘埃和分子云的分布，而X射线图象显示了高温（千万度）热气体的分布和其中的所存在的中子星。要研究这类天体的物理过程，就必须结合几个波段上的数据来一起进行分析。现在，我们已经拥有了十个波段上的巡天数据，不久就会再增加五个以上的波段巡天。

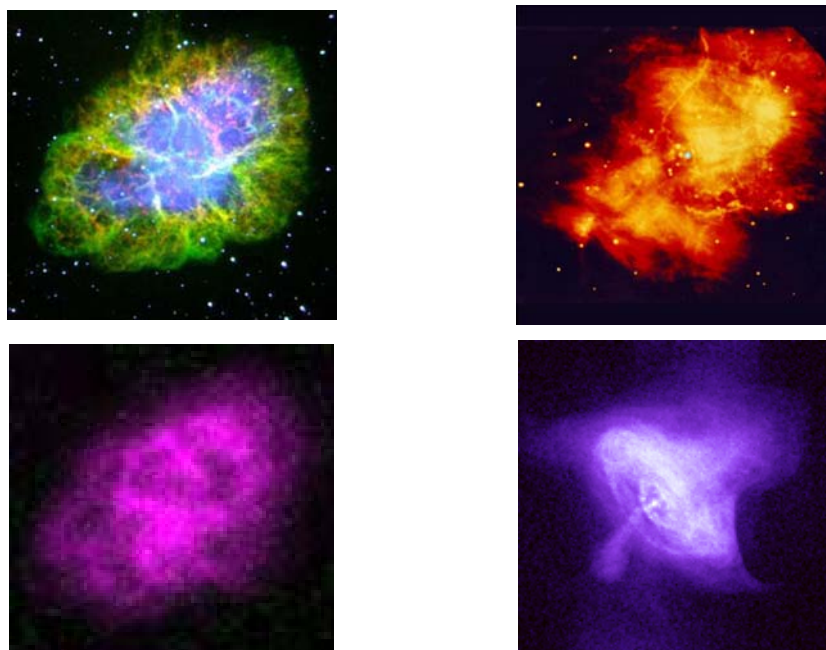


图 1.1 蟹状星云的多波段图象（左上：光学、左下：射电、右上：红外、右下：X 射线）。蟹状星云是 1054 年中国天文学家发现的一颗超新星爆发后所留下的遗迹。

3. 天文数据量正在呈指数级增长，而全世界的天文数据量已经达到PB（ $10^{15}$  Byte）量级，目前天文数据翻番的时间已经不到一年。其直接原因如下：
  - (1) 一大批天文观测设备的投入运行，特别是许多数字巡天项目的实施，带来了海量天文数据的产出；
  - (2) 天文数值模拟也产生的大量的模拟数据；

- (3) 对天文数据的深加工也产生了大量的数据产品;
- (4) 对过去资料的数字化也加速了天文数据的增长;
- 4. 天文数据结构和格式的复杂度也随着天文数据的增长而增加;
- 5. 大部分天文数据都经过良好组织和归档, 当前世界上有许多家天文数据中心在数据的整理归档整理发布方面做了大量的工作。用于天文数据处理的软件包也日益变得功能更强大, 种类更丰富和多样化。

## § 1.2 虚拟天文台的概念<sup>[1,13]</sup>

数百年来, 天文研究通常都是单个或几个天文学家进行少数天体的观测, 加上大望远镜的观测时间非常有限, 就无法研究那些需要大量数据来解决的天体物理问题。而在今后, 技术上的进步将导致前所未有的天文数据流, 造成所谓的“数据雪崩”现象。与以往不同的是, 这些数据大部分是质量均匀的、常常在多个波段上并覆盖相当大的天区, 其科学内容的丰富和深度是空前的, 因而能为天文学家提供前所未有的机会来利用这些数据进行各种科学问题的研究。

面对这些包含数亿个天体的多波段的海量数据, 天文学家显然需要访问这些数据和分析其中数据的工具, 来进行诸如数据挖掘、先进的模式识别、大规模统计交叉关联、罕见天体以及时变的发现等工作。此外, 有了这样的海量数据, 在天文学历史上第一次可以将复杂的数值模拟与统计上完备的数据进行有意义的对比。而迅速发展的高速网络意味着全世界的天文学家都可以分享这些科学成果。

如果说利用伽马射线巡天、X 射线巡天、紫外巡天、光学巡天、红外巡天和射电巡天所得到的观测数据, 用适合的方法对数据进行统一规范的整理、归档, 便可以构成一个全波段的数字虚拟天空; 而根据用户要求获得某个天区的各类数据, 就仿佛是在使用一架虚拟的天文望远镜; 如果再根据科学研究的要求开发出功能强大的计算工具、统计分析工具和数据挖掘工具, 这就相当于拥有了虚拟的各种探测设备。这样, 由虚拟的数字天空、虚拟的望远镜和虚拟的探测设备所组成的机构便是一个独一无二的“虚拟”天文台。由此可见, 虚拟天文台是在互联网时代里天文学发展的必然产物。

新的巡天带来了巨大的科学发现的潜力，对这些巡天数据的联合使用，将涌现出全新的、无法预见的、意义重大的科学产出，这是一种仅靠单独使用其中某一部分数据所不能产生的新科学。人们越来越意识到，科学数据的获得、组织、分析和传播是持续而坚实地发展科学技术的基本要素，所有这些都表明，建立一种结构去有效地综合这些技术能力势在必行。因此现在就需要一个像虚拟天文台这样的实体来指导日益增多的天文数据的合理部署。所以有必要投入一定的人力、财力、物力将所有符合特定规范的数据联合到虚拟天文台中，其科学意义是传统天文台所无法比拟也无法替代的。

虚拟天文台将使天文学取得前所未有的进展，它将成为开创“天文学发现新时代”的关键性因素。虚拟天文台将是独一无二的，它将万亿字节的数据库、波长遍及从伽玛射线到射电波段的数亿个天体的图象库、高度复杂的数据挖掘和分析工具、具有千万亿字节容量的存储设备和每秒运算次数达到万亿次的超级计算设备、以及各主要天文数据中心之间的高速网络连成一体；它使世界各地的天文学家可以快速查询各个万亿字节大小的数据库；使埋藏在庞大星表和图象数据库中的多变量模式可视化；增加发现复杂规律和稀有天体的机会；鼓励多个研究团体的实时合作；允许进行大规模的统计研究，将首次使数据库的内容可以和复杂精密的数值模拟结果进行对比。虚拟天文台将促进我们对许多决定宇宙演化的天体物理过程的理解，它会用更经济的投资产生新的和更好的科学。虚拟天文台将作为一个协调性的和操作性的机构来促进新型的工具、协议和合作方面的发展，以充分实现未来十年内天文数据库的科学潜能，虚拟天文台将成为“天文学发现”的推进器。

目前，天文学家确定的虚拟天文台的主要科学目标是<sup>[1,14]</sup>：

1. 多观测参数空间的探索：将各个巡天数据统一到虚拟天文台中，将会有更广泛而复杂的应用。这些数据能提供全天在十多个不同波段上的信息，在多维空间里展示整个天空的真实面貌。可以说，多种巡天数据在虚拟天文台中的完美结合，将会得到更加完善的真实的宇宙图像（多层次的、大尺度的、系统性的等等）。

2. 稀有天体与新型天体的发现：目前通过巡天来寻找稀有天体（如高红移类星体、褐矮星等）的项目正在蓬勃发展。假如某种有趣的天体或现象出现的概



率是百万分之一或一亿分之一，那么就需要几百万或几亿个样本才有可能发现，这样，在海量数据中进行彻底的宇宙探索来寻找稀有的未知类型天体具有更加诱人的前景。因此，虚拟天文台将会促进新的天文发现。

3. 新兴的科学领域：虚拟天文台对任何要求融合各类数据来研究天文现象的课题都具有重要的影响。虚拟天文台的出现促进了多波段天文学的发展，不同波段的巡天数据的联合可以从更深层次来探索宇宙；同时，虚拟天文台推动了各种各样令人兴奋的科学探索，如活动星系核和星系团的多层次研究、低表面亮度星系的形成和演化的研究、星系结构的研究等；虚拟天文台的出现还促进了统计天文学的兴起，如宇宙大尺度结构和银河系结构的定量分析、各种天体（特殊种类或特殊性质的恒星或星系、活动星系核、星系团等）完备样本的建立与研究，等等。虚拟天文台的建立可以使天文学研究在数量和质量上充分地提高。

4. 天文学中的数据挖掘：从海量数据中发现稀有的天体或现象，或者发现以前未知种类的天体或新的天文现象，或者根据数据来区分不同类型的天体等，都需要充分运用在信息科学中迅速发展的数据挖掘和知识发现技术。数据挖掘技术在虚拟天文台中的应用，将使任何地方的天文学家在不依赖于大望远镜的情况下就可以做出一流的工作，而这种研究方式完全不同于传统的天文学研究。运用数据挖掘技术可以有效地解决天文学中的“数据雪崩”问题，这对天文学发展是至关重要的<sup>[15]</sup>。

### § 1.3 虚拟天文台的功能和需求<sup>[2]</sup>

#### 1. 资源的发现机制

虚拟天文台是一个动态的环境，新的数据资源产生，原有的数据资源在更新，一些旧的数据资源被舍弃，虚拟天文台必须提供动态发现数据资源和描述数据资源属性和包括什么样的数据的机制，每一个数据资源在由观测天区范围、观测波段、分辨率和观测历元等属性组成的多维参数空间里都是一个多方体，所以这些属性信息应该被收集并提供到实现数据发现机制中去。这样天文学家就可以去通过这个发现机制来查找他所感兴趣的数据资源。

虚拟天文台不仅仅能发现数据资源，而且为了天文学中海量计算的需求，还

应该能提供发现计算资源的机制, 这样通过利用分布式的计算资源可以完成以前被认为是不可能的, 或代价昂贵的计算处理要求。由于天文中大部分海量计算处理伴随着海量的数据, 所以还需要虚拟天文台在发现这些可用的分布式计算资源的计算能力的同时, 还能提供描述这些计算资源的存储能力, 网络带宽、计算环境的信息。

资源的概念远远不止数据资源和计算资源, 还包括其他天文学家需要的其他资源, 包括文献资源, 数字图书资源等等。所以虚拟天文台也有必要提供这些资源的发现功能机制。

## 2. 提供无缝获取全球的分布式数据的功能

目前天文数据被分布存储在全球上千个地点的数据中心内, 将这些海量数据集中在一个地点存储显然是不现实的, 为了使天文学家和计算机程序能使用到这些分布式的数据, 这些数据必须符合一定的元数据标准。虚拟天文台的一个挑战是扩展天文数据的语义标准, 使得能通过一般术语描述查询数据的要求, 并且查询结果能够被计算机和人都能理解。

虚拟天文台的一个重要的目标就是实现分布式数据的统一访问机制, 它能为天文学家提供一个统一的访问这些分布数据接口, 并能为天文学家提供描述查询要求和获取查询结果的便利。这个访问机制不仅仅提供给天文学家, 这个机制还提供给用户应用程序, 使得程序也可以通过一个统一的访问接口来查询获取数据。

## 3. 融合全球的分布式数据

数据融合是指将多个独立的星表, 数据库或数据集整合成一个整体, 融合现有的数据集也能产生新的发现。一个融合星表的重要工作就是基于不同星表内的天体在位置上相关性, 进行星表间的交叉证认工作。

## 4. 改造现有的数据管理系统, 开发新的数据归档标准

采用数据库管理系统来组织、管理和获取天文数据在近 10 年来得到显著的增长, 增长的驱动力来自于数据量的飞速增长, 对复杂查询需求的增长以及对低成本商业关系型数据库管理软件的采用。但是目前的数据库管理系统并不能满足未来虚拟天文台海量数据的管理需要, 虚拟天文台为未来的数据库管理系统提出了如下需求, 有必要对目前数据库系统进行优化改造, 使之能满足未来需要。

- (1) 提供数据元数据的高效管理；
- (2) 具备高速的数据导入数据库的能力，现在数据库的数据导入能力成为了一个管理海量数据的瓶颈；
- (3) 提供对复杂结构数据的高效管理。我们知道关系型数据库在管理结构化数据方面具有很强的能力，但是在管理非结构化数据比如 XML 方面功能较差。
- (4) 提供对海量数据的高效管理，目前的数据库管理技术已经不能满足管理海量数据的需要。比如像结构最简单的星表，目前最大的星表包括约 10 亿颗天体的信息，采用数据库管理存储的数据量将会达到 160GB，查询这样大星表的效率已经非常低下，不久将会有数据量大于 1TB 的星表出现<sup>[16]</sup>，管理这样的星表更是目前数据库系统需要解决的问题。而像管理图像等更复杂、数据量更大（PB 量级）的数据的难度则更大。
- (5) 提供高效的复杂查询功能，为用户提供天文专用的查询功能。实现对海量数据的高速查询需要数据库系统提供有效的索引，完成用户复杂的查询要求，特别是天文上一些专用的查询要求，不但需要高效的索引，还要数据库管理系统提供功能强大的函数，过程和事务处理功能。
- (6) 优良的结构框架和接口为基于数据库系统的数据融合和数据挖掘提供便利。

虚拟天文台是基于现有的数据库来提供天文数据的获取的，所以需要接口化现有的商业和非商业的数据库管理系统，元数据格式和结构，以及能实现跨不同数据集的复杂查询功能的查询协议。实现接口化需要制订一系列的标准，这些标准包括数据组织、元数据格式和内容，查询协议接口，以及数据交换格式和协议等等，这些将导致包括一个处在查询协议和管理天文数据的数据库系统之间的翻译层的虚拟天文台框架结构，这个翻译层将使得将目前的天文数据融入到虚拟天文台中去所需要做的工作最小化。这些协议的制定也是迫切需要解决的，因为这些标准的制定出台使得将来新的数据集的设计和生成有规可循，并且能充分利用虚拟天文台提供的便利，用最小的成本实现最大的功能，降低数据发布的成本。

## 5. 开发高效处理海量数据的处理软件

对海量数据的处理必须要有高效的处理软件，目前的数据处理软件不能满足

快速处理的要求。还有一个重要的是建立完善的软件工具集，并实现软件代码的高度可复用性。

#### 6. 发展可视化海量数据的技术和工具

虚拟天文台为天文学家提供了海量的数据信息，如何更加直观的对这些数据进行探察，需要为天文学家开发针对于星表、图像和数值模拟数据的可视化的技术和工具，虚拟天文台的用户们会经常处理  $N$ -维参数空间，所以希望通过可视化理解参数间的相关性和找到离群数据。

#### 7. 开发数据挖掘和知识发现的新算法和工具<sup>[17]</sup>

数据挖掘包括在数据集中寻找模式，并能在其中证认出非正常事件，这些有可能是很稀有的现象或者是数据的缺陷。虚拟天文台必须提供运行于数据融合基础之上的回归分析、偏差分析、序列分析、通过聚类和相似性的分类的算法。新的数据挖掘工具必须融入到虚拟天文台中去，并且能够充分利用优化后的数据库结构和丰富的元数据信息带来的便利。

还需要开发能从数据挖掘的结果中导出新知识的知识发现的新工具。知识发现一个重要的作用就是实现天文学家和梗概数据（从数据挖掘分析中得到的结果）可交互性，使得天文学家直接从数据中寻找知识的过程更加便利，并能产生出科学结果——文字、图形或发表的文献。这是一个相当艰难的任务。

### § 1.4 国际上虚拟天文台的发展现状

1999 年美国天文学家首先提出了“虚拟天文台”的计划，其目的是作为促进和协调的实体，为实现海量天文数据的全部科学潜在价值，实现数据的高度共享和融合，发展各种必要的手段工具，制定协议，组织合作。美国国家科学院天文学和天体物理学发展规划委员会在题为“新千年天文学和天体物理学”的未来十年发展规划中把建立国家虚拟天文台（NVO<sup>[18]</sup>）做为最优先的推荐项目。2003 年 NVO 已经获得国家科学基金立项，获得 5 年的资金资助 1000 万美元。

虚拟天文台的计划在各国天文界引起了广泛的反响，许多国家纷纷建立自己的虚拟天文台计划，在全球掀起了一场虚拟天文台的研究浪潮，中国也是其中一员。当前国际上主要的虚拟天文台研究项目如下表所示：

项目	地区	网址
中国虚拟天文台 (China-VO <sup>[19]</sup> )	中国	<a href="http://www.china-vo.org/">http://www.china-vo.org/</a>
国家虚拟天文台 (NVO)	美国	<a href="http://www.us-vo.org/">http://www.us-vo.org/</a>
英国虚拟天文台 (AstroGrid <sup>[20]</sup> )	英国	<a href="http://www.astrogrid.org/">http://www.astrogrid.org/</a>
欧洲虚拟天文台 (Euro-VO、AVO <sup>[21]</sup> )	欧洲	<a href="http://www.euro-vo.org/">http://www.euro-vo.org/</a>
法国虚拟天文台 (VO-France <sup>[22]</sup> )	法国	
加拿大虚拟天文台 (CVO <sup>[23]</sup> )	加拿大	<a href="http://services.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/cvo/">http://services.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/cvo/</a>
印度虚拟天文台 (VO-India <sup>[24]</sup> )	印度	<a href="http://vo.iucaa.ernet.in/~voi/">http://vo.iucaa.ernet.in/~voi/</a>
德国天体物理虚拟天文台 (GAVO <sup>[25]</sup> )	德国	<a href="http://www.g-vo.org/">http://www.g-vo.org/</a>
澳大利亚虚拟天文台 (Aus-VO <sup>[26]</sup> )	澳大利亚	<a href="http://www.aus-vo.org/">http://www.aus-vo.org/</a>
日本虚拟天文台 (JVO <sup>[27]</sup> )	日本	<a href="http://jvo.nao.ac.jp/">http://jvo.nao.ac.jp/</a>
俄罗斯虚拟天文台 (RVO <sup>[28]</sup> )	俄罗斯	<a href="http://www.inasan.rssi.ru/eng/rvo/">http://www.inasan.rssi.ru/eng/rvo/</a>
意大利虚拟天文台 (DRACO <sup>[29]</sup> )	意大利	<a href="http://www.oat.ts.astro.it/draco/DRACO-home.htm">http://www.oat.ts.astro.it/draco/DRACO-home.htm</a>
韩国虚拟天文台 (KVO <sup>[30]</sup> )	韩国	<a href="http://kvo.kao.re.kr/">http://kvo.kao.re.kr/</a>
匈牙利虚拟天文台 (HVO <sup>[31]</sup> )	匈牙利	<a href="http://www.h-vo.org/">http://www.h-vo.org/</a>

表 1.1 国际上的虚拟天文台研究项目计划

虽然这些计划来自不同的国家，有着不同的天文和信息技术背景，但是他们之间有许多非常重要的共同点：每个项目都在寻求数据密集型天文研究的出路，都在力图挖掘现有以及未来海量天文数据的潜力<sup>[4]</sup>。

每个虚拟天文台的项目都制定了有自己特色的科学目标和实施计划，都有着自己的技术优势和兴趣。但虚拟天文台都面临着共同的需求并寻求着相似的解决方案，从整个天文界的眼光来看，虚拟天文台并不局限于每个项目所属的国家内部，而是建立全球范围国际虚拟天文台 (IVO)。

因为IVO必须是一个完整的，不同的部分之间能进行互操作的系统。为了实现不同部分之间的互操作性，就必须对许多问题标准化在国际范围内达成一致。

这其中最重要的就是数据和接口的标准化，此外需要协调的还包括软件包、源代码库、开发工具等等。而且各虚拟天文台之间的密切合作和资源互补也是非常重要的。正是出于此目的，AstroGrid、AVO、NVO 联合在德国Garching 举行的“迈向国际虚拟天文台<sup>[32]</sup>（Towards an International Virtual Observatory）”国际会议上提出组建国际虚拟天文台联盟（IVOA<sup>[33]</sup>）的倡议，阐述了IVOA的历史使命并提出了相应的行动方案。其使命是：推进国际合作与协作，为建设一个能综合利用全球天文数据的、完整的、能协同工作的虚拟天文台，开发、配置必要的工具、系统和组织结构。

从 2002 年 6 月成立之初的 8 个成员，到 2004 年 5 月 IVOA 已经发展到包括上表所示 14 个成员。IVOA 到现在成立已经近两年，在这两年里，IVOA 促进了各个虚拟天文台项目间的合作与沟通，实现了人力、物力和财力的高效利用，扩大了虚拟天文台在国际的声誉和影响，加速了国际虚拟天文台的建设步伐。



图 1.2: 国际虚拟天文台联盟全家福<sup>[34]</sup>

## § 1.5 中国虚拟天文台 (China-VO)

1999 年美国天文学家首先提出了“虚拟天文台”的概念后,中国的天文学家也认识到了虚拟天文台在未来的天文学发展中具有重大意义,在 2001 年 9 月召开了“虚拟天文台学术研讨会”,并成立了我国自己的“虚拟天文台科学工作组”,这标志着中国虚拟天文台 (China-VO) 的成立。

中国虚拟天文台的建设具有非常重要的意义<sup>[1,35,36]</sup>:

1. 与在信息技术领域的互联网建设相似,中国天文界只有在虚拟天文台的建设中作出自己的贡献,才会以平等的身份全方位共享虚拟天文台的技术与资源。

2. 中国虚拟天文台整合国内国际天文资源,服务主要面向国内的天文学家,为天文学家提供一个全新的研究工作平台,为中国天文学的更大发展提供基础支持和契机。虚拟天文台对于我国这样的发展中国家来说具有更加重要的意义,国力的限制使我们不可能发展各类的大型天文观测设备,而由于国际上天文数据的开放性和虚拟天文台的建设与发展,我国的天文学家完全可以充分利用国际上最先进的设备所获得的高质量数据来做出一流的科学研究工作,提升中国天文学家在国际天文学界的地位,虚拟天文台是中国天文学新的发展机遇。

3. 中国天文界一直以来缺乏一个整理组织自己数据的数据中心,缺乏数据的共享实际上是巨大的资源浪费,虚拟天文台的一个重要作用是提供一个新的更加高效和经济的数据发布机制,并提供相关的工具和标准,中国虚拟天文台将会为 LAMOST 的数据产品的发布提供极大的便利,并且将会以发布 LAMOST 数据为契机,承担起整个中国天文数据的整理和发布。

4. 虚拟天文台架起了天文学与计算机科学之间的桥梁,为我国天文学家与计算机学家的精诚合作提供了良好的机遇,为国内的计算技术、网格技术、数据库技术和数据挖掘技术的发展提供一个良好的实验研究平台,促进了计算机科学的发展。

5. 建设虚拟天文台可以培养一批与虚拟天文台相适应的天文学家和技术人才,为未来中国天文学发展提供智力支持;

6. 由于大量真正的科学数据能够免费地从因特网上获得,而且公众对天文

学有着浓厚的兴趣，虚拟天文台将特别适合教育和科学普及。虚拟天文台具有基于因特网的内禀性质，因此能够在前所未有的社会和地理范围内提供各式各样的高质量科普和教育方法，提高我国公众的科学素质。

我们相信，虚拟天文台是当代天文学中发展新科学的推进器、产生新发现的催化剂，它将开辟出探索宇宙的新天地。在这历史机遇面前，中国科学家也应该也一定能够为虚拟天文台的发展作出自己应有的贡献。



## 参考文献

- [1] 赵永恒. 互联网时代的天文学革命-虚拟天文台. 科学, 2002, 54(2):13
- [2] Towards a National Virtual Observatory: Science Goals, Technical Challenges, and Implementation Plan, June 2000(white paper).  
[http://www.astro.caltech.edu/nvoconf/white\\_paper.pdf](http://www.astro.caltech.edu/nvoconf/white_paper.pdf)
- [3] Paul Messina, Alex Szalay. Project Description: Building the Framework for the National Virtual Observatory. <http://bill.cacr.caltech.edu/cfdocs/usvopubs/files/nvo-proj.pdf>
- [4] 崔辰州. 博士学位论文 中国虚拟天文台系统设计
- [5] 程景全. 天文望远镜原理和设计——射电、红外、光学、X 射线和射线望远镜. 第 1 版. 北京: 中国科学技术出版社, 2003. 325
- [6] [OWL]. The ESO 100-m OWL optical telescope concept.  
<http://www.eso.org/projects/owl/>
- [7] [SKA] The Square Kilometre Array. <http://www.atnf.csiro.au/projects/ska/>
- [8] [SDSS] Sloan Digital Sky Survey <http://www.sdss.org/>
- [9] [LAMOST] The Large Sky Area Multi-Object Fiber Spectroscopic Telescope.  
<http://www.lamost.org/>
- [10] [LSST] Large-aperture Synoptic Survey Telescope. <http://www.lsst.org/>
- [11] Andrew Connolly University of Pittsburgh. LSST Data Management: Prospects for Processing and Archiving Massive Astronomical Data Sets.  
[www.lsst.org/Project/docs/lsst\\_data\\_man\\_prospects.pdf](http://www.lsst.org/Project/docs/lsst_data_man_prospects.pdf)
- [12] Szalay A, Gray J. The World-Wide Telescope. Science. 2001(293):203
- [13] 陈建生, 赵永恒. 虚拟天文台-新世纪的天文学革命. 虚拟天文台研讨会, 2001, Sep
- [14] 张彦霞, 赵永恒, 崔辰州 天文学进展 2002,20(4):312
- [15] Banaday A., Zaroubi S, Barteman M. Mining the Sky, Springer-Verlag, Berlin, 2001
- [16] LSST Phys Colloquium. <http://www.lsst.org/Project/docs/LSST-SW.ppt>

- [17] Peter Allan,Bob Bentley. AstroGrid THE SCIENCE PROGRAMME .  
<http://www.pparc.ac.uk/Pbl/Report2001Programme.pdf>
- [18] [NVO] US National Virtual Observatory. <http://www.us-vo.org>
- [19] [China-VO] Chinese Virtual Observatory.<http://www.china-vo.org>
- [20] [AstroGrid] VO United Kingdom. <http://www.astrogrid.org>
- [21] [AVO] Astrophysics Virtual Observatory. <http://www.euro-vo.org>
- [22] [VO-France] France Virtual Observatory. <http://www.france-vo.org>
- [23] [CVO] Canadian Virtual Observatory.  
<http://services.cadc-ccda.hia-ihp.nrcnrc.gc.ca/cvo/>
- [24] [VO-India] Virtual Observatory India. <http://vo.iucaa.ernet.in/~voi/>
- [25] [GAVO] German Astrophysical Virtual Observatory. <http://www.g-vo.org/>
- [26] [Aus-VO] Australian Virtual Observatory. <http://www.aus-vo.org>
- [27] [JVO] Japanese Virtual Observatory. <http://jvo.nao.ac.jp/>
- [28] [RVO] Russian Virtual Observatory. <http://www.inasan.rssi.ru/eng/rvo/>
- [29] [DRACO] Datagrid for Italian Research in Astrophysics and Coordination with the Virtual Observatory. <http://www.as.oat.ts.astro.it/draco/>
- [30] [KVO] Korean Virtual Observatory. <http://kvo.kao.re.kr/>
- [31] [HVO] Hungarian Virtual Observatory. <http://hvo.elte.hu/en/>
- [32] [Garching2002] Toward an International Virtual Observatory: Scientific Motivation, Roadmap for Development and Current Status.  
<http://www.eso.org/gen-fac/meetings/vo2002/>
- [33] [IVOA] International Virtual Observatory Alliance. <http://www.ivoa.net>
- [34] IVOA Member Organizations. <http://www.ivoa.net/pub/members/index.html>
- [35] 张彦霞,赵永恒. 虚拟天文台. [http://www.china-vo/docs/VO\\_zyx.PDF](http://www.china-vo/docs/VO_zyx.PDF)
- [36] 赵永恒,崔辰州. 中国虚拟天文台——任务、特点、方案.  
[http://www.china-vo.org/docs/cvo\\_draft.pdf](http://www.china-vo.org/docs/cvo_draft.pdf)



## 第二章 VO 数据标准

### § 2.1 VO 数据标准简介

为了实现虚拟天文台提出的功能，实现数据资源的互操作，必须制订和采纳一系列的虚拟天文台标准，目前各虚拟天文台都在积极地进行标准的研究和制订，IVOA 目前成立了多个工作组，统筹天文各方面的数据标准研究修订和应用工具的开发工作。目前比较热门的工作讨论组以及他们涉及的标准如下：

工作组和论坛	主题	涉及的标准
<a href="mailto:dal@ivoa.net">dal@ivoa.net</a>	数据访问层	
<a href="mailto:dm@ivoa.net">dm@ivoa.net</a>	数据模型	数据模型标准
<a href="mailto:grid@ivoa.net">grid@ivoa.net</a>	Grid&Web 服务	服务标准
<a href="mailto:registry@ivoa.net">registry@ivoa.net</a>	元数据资源/注册	资源元数据标准
<a href="mailto:stdproc@ivoa.net">stdproc@ivoa.net</a>	标准化处理	
<a href="mailto:ucd@ivoa.net">ucd@ivoa.net</a>	UCD (统一内容描述)	元数据
<a href="mailto:voql@ivoa.net">voql@ivoa.net</a>	VOQL (虚拟天文台查询语言)	查询语言标准
<a href="mailto:votable@ivoa.net">votable@ivoa.net</a>	VOTable/XML	数据封装标准
<a href="mailto:architecture@ivoa.net">architecture@ivoa.net</a>	系统体系结构	体系结构标准
<a href="mailto:apps@ivoa.net">apps@ivoa.net</a>	应用	
<a href="mailto:net@ivoa.net">net@ivoa.net</a>	网络讨论	
<a href="mailto:radiovo@ivoa.net">radiovo@ivoa.net</a>	射电天文	
<a href="mailto:semantics@ivoa.net">semantics@ivoa.net</a>	语义, 知识工程	
<a href="mailto:theory@ivoa.net">theory@ivoa.net</a>	VO 理论	
<a href="mailto:interop@ivoa.net">interop@ivoa.net</a>	互操作性	
<a href="mailto:ivoa@ivoa.net">ivoa@ivoa.net</a>	IVOA 执行委员会	

表 2.1: IVOA工作组和涉及的标准<sup>[1]</sup>

作为虚拟天文台大家庭的一员，中国虚拟天文台也应该积极参与到相关标准的制订修改和采纳应用中来。采纳应用相关的标准才能实现虚拟天文台的功能，并保持与国际虚拟天文台的一致性和互操作性，而且只有将标准投入应用，才能发现其中的不足和缺陷，为标准的修订提供实践经验和依据，指导修订的方向。另外，我们还需要参与到标准的推出和修订中来，应该认识到这也是非常重要的，原因是我们在开发和标准的应用中必定会发现对新标准的需求和现有标准的缺陷，这些将会阻碍应用的实现，必须要参与标准的推出和修改，静观其变肯定不是个好主意。并且在实践中得出的这些意见也是很宝贵的，应该与同行共享这些意见。

在实现中国虚拟天文台的基于网格环境的数据获取服务中，我们采纳了一系列的数据标准，这些标准包括以下部分，如果缺乏这些数据标准的支持，就很难实现天文数据的统一获取和互操作。

1. VOTable，作为天文星表数据的标准封装格式；
2. 统一内容描述符 UCD，已经被包括到 VOTable 格式中，用来作为列的语义元数据；
3. 天文数据查询语言-ADQL，作为描述查询星表数据库要求的标准语言；
4. 天文资源元数据标准（Resource Metadata Standard），作为数据资源的元数据描述标准被应用到对数据服务资源的登记注册上。

## § 2.2 天文表列数据的标准封装格式——VOTable

### 2.2.1 VOTable 概述

VOTable<sup>[2]</sup>是天文界为了满足数据互操作性的需要而制定的基于XML语言的表数据集合的格式规范，用做灵活交换传输和存储表列结构数据的，特别是作为天文表列数据的标准数据格式。VOTable格式前身是Astros<sup>[3]</sup>格式，VOTable格式结构接近于FITS二进制表(FITS Binary Table)的格式标准。VOTable采用XML格式标准使得应用很容易验证VOTable的有效性，并能很容易地转换（利用XSLT<sup>[4]</sup>转换器）成其他数据格式。

VOTable 的一个重要特点是提供了对表的元数据的描述定义。VOTable 不仅仅

包括表数据，而且还包括对描述这些表数据的元数据。元数据包括对表自身的描述和对表中各属性列的描述，在对属性列的描述中还引入了描述列的具体天文语义的UCD，使得计算机和人都可以理解每一列数据表达的具体天文意义。

**VOTable格式能够满足大数据量网格计算的需要**<sup>[5,6,7,8,9]</sup>。VOTable格式允许数据和元数据分离存储，即在文档中仅仅包含指向远程数据的链接(而不包含数据本身)和元数据信息。在网格计算中，处理器可以根据这些元数据信息为接受远程数据做好准备，或者是组织第三方传输或并行传输远程的数据。当我们在网格分布式计算环境中运用大数据量的表列数据时，数据流在异地的处理器间传输，并被过滤和缓存，如果需要在数据表的头部提供数据的行数的话（比如FITS）会带来很大的麻烦和困难，必须要把整个数据表输入流缓存在存储器中，计算有多少行，然后把数据再次转换成数据流，要知道，在网格中最缺乏的不是处理器，而是大的缓冲存储器，而VOTable不需要在文件头提供数据的大小。另外，由于VOTable可能是由其他的处理器动态生成或存储在一个暂时的存储器中，VOTable可以提供一个参数描述经过多久远程数据将不再存在。在网络上的天文数据有可能需要获取授权，所以VOTable也提供描述了密码和身份确认信息的参数。

**VOTable 是灵活高效的数据存储格式。**VOTable 的数据部分可以选择采用三种不同的存储格式，包括 TABLEDATA, FITS, BINARY。TABLEDATA 是一个纯 XML 的格式，所以能采用常规的 XML 工具对小数据表高效处理；FITS Binary Table 是天文学家最熟悉的格式，VOTable 能作为数据传输和存储的标准格式，也能作为单独存储元数据的标准格式。VOTable 可以被用来将整个 FITS 文件封装，或重新对元数据进行重新编码；很不幸的是最困难的是流化 FITS 文件，因为 FITS 头需要数据的长度值，并且需要确定变长数组的最大长度；BINARY 格式即采用二进制存储数据，采用这种格式节省存储空间并且容易编写程序，在存储空间的利用上甚至强于 FITS（FITS 文档的一个记录长度必须是 2880 字节，会浪费空间），并且读写 BINARY 数据也不需要专门的 FITS 的程序库。

### 2.2.2 VOTable 的数据模式

<b>VOTable</b>	=	hierarchy of Metadata + associated Data
<b>Metadata</b>	=	Parameters + Infos + Descriptions + Links + Fields + Groups
<b>Table</b>	=	list of Fields + TableData
<b>TableData</b>	=	stream of Rows
<b>Row</b>	=	list of Cells
		Primitive
<b>Cell</b>	=	or variable-length list of Primitives
		or multidimensional array of Primitives
<b>Primitive</b>	=	integer, character, float, floatComplex, etc (see the table 1 below).

表 2.2: VOTable的数据模式<sup>[5]</sup>

VOTable 的数据模式见表，VOTable 由表列数据 (Data) 和描述这些数据的元数据 (Metadata) 组成。元数据包括两部分，一是表自身信息的元数据 (Parameters, Infos, Descriptions)，另一是表数据的每一列的元数据信息 (包括 Fields 和 Groups)。Fields 包含若干个 FIELD 元素，是对每一列数据元素进行描述，FIELD 个数与数据列数要严格一致，每一个 FIELD 元素包含若干属性和子元素，描述对应列的名称 (name 属性)，ID (ID 属性)，UCD (ucd 属性)、数据类型 (datatype 属性)、宽度 (width 属性)、精度 (precision 属性)、数组各维长度 (arraysize 属性) 和详细注释 (DESCRIPTION 元素)，如果某一 FIELD 元素 name 属性为 RA，那么就表示对应的一列列名为 RA。

PARAM 元素结构与 FIELD 基本相同，只是多了一个 value 属性，它可以被用来表示常量数据，例如可以表示 FITS 中的 Keyword，或者表示这个产生这个表的相关参数信息。例如在星表查询中，可以将查询条件包含在 PARAM 中。

从 VOTable1.1 版本开始，列可以构成逻辑组 (GROUP)，这个组一般包括多个列 (FIELD) 和相关的参数(PARAM)。组可以将一系列有相互关系的列逻辑上联合起来，从而表示列 (FIELD) 之间以及与参数(PARAM)的关联关系

表数据 (TableData) 是文本格式的，用来存储表列数据，由若干行 (Rows) 组成，而行由若干 (cell) 组成，而由一个、变长多个或多维的原生数据类型 (Primitive datatype) 的数据组成，VOTable 支持的原生数据类型如下表所示：

<b>datatype</b>	Meaning	<b>FITS</b>	Bytes
"boolean"	Logical	"L"	1
"bit"	Bit	"X"	*
"unsignedByte"	Byte (0 to 255)	"B"	1
"short"	Short Integer	"I"	2
"int"	Integer	"J"	4
"long"	Long integer	"K"	8
"char"	ASCII Character	"A"	1
"unicodeChar"	Unicode Character		2
"float"	Floating point	"E"	4
"double"	Double	"D"	8
"floatComplex"	Float Complex	"C"	8
"doubleComplex"	Double Complex	"M"	16

表 2.3: VOTable 支持的原生数据类型<sup>[5]</sup>

### 2.2.3 VOTable 格式与 FITS Binary Table 格式的兼容性<sup>[5]</sup>:

VOTable 格式接近与兼容 FITS Binary Table 格式，在此我们简称 FITS Binary Table 为 FITS，一个 FITS 文件如何转换成一个 VOTable 呢？可以将 FITS 头转换成 VOTable 格式文件，可以使用 VOTable 中的 PARAM 元素来表示 FITS 中的 Keyword 这些名/值序列，然后在这个文件中添加指向这个文件的链结。不过需要提出的是，目前没有具体的规定如何将 FITS 转换成 VOTable 文档，可以采取的方式很多，这或许会在今后的使用中带来混乱和麻烦。

#### 什么是 FITS 能做而 VOTable 不能做的：

FITS 提供了将一组字符串构造成一个字符串的功能（即字符串数组），而 VOTable 1.0 和 1.1 版本不支持这项功能。

#### 什么是 VOTable 能做而 FITS 不能做的：

VOTable 支持元数据和数据的分离存储，即将元数据存储于 VOTable 文档中，而数据存在于 FITS 中，在 VOTable 建立指向这个 FITS 文档的链结，因此也架起了 XML 和 FITS 格式的桥梁。VOTable 通过 UCD 描述数据列和参数的语意内容，这是 FITS 做不到的。VOTable 还支持 FITS 不支持的 Unicode 字符。

另需要说明的是，从 FITS 转换成 VOTable 是不可逆的，从 FITS 转换成 VOTable 不会丢失任何信息，但是相反的反转换不丢失信息是不可能的。



图 2.1 是一个 VOTable 文档的例子：

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="http://www.lamost.org/~sang/VOTable-basic.xsl"?>
<!DOCTYPE VOTABLE SYSTEM "http://www.lamost.org/~sang/VOTable.dtd">
<VOTABLE version="1.0">
  <DESCRIPTION>This is an example VOTable document</DESCRIPTION>
  <DEFINITIONS>
    <COOSYS ID="myJ2000" equinox="2000." epoch="2000" system="eq_FK5"/>
  </DEFINITIONS>
  <RESOURCE name="GSC1.2">
    <DESCRIPTION>
      This is an excerpt of the HST Guide Star Catalog, Version 1.2 (Lasker+ 1996).
    </DESCRIPTION>
    <TABLE>
      <FIELD ID="ra" name="RA(J2000)" ref="myJ2000" ucd="POS_EQ_RA_MAIN"
        unit="deg" datatype="double" precision="F5">
        <DESCRIPTION>Right ascension in J2000, epoch of plate</DESCRIPTION>
      </FIELD>
      <FIELD ID="dec" name="Dec(J2000)" ref="myJ2000" ucd="POS_EQ_DEC_MAIN"
        unit="deg" datatype="double" precision="F5">
        <DESCRIPTION>Declination in J2000, epoch of plate</DESCRIPTION>
      </FIELD>
      <FIELD ID="mag" name="Pmag" ucd="PHOT_PHG_MAG" unit="mag"
        datatype="float" width="5" precision="2">
        <DESCRIPTION>photographic magnitude (see n_Pmag)</DESCRIPTION>
      </FIELD>
      <LINK content-role="doc" title="doc" href="http://vizier.u-strasbg.fr/viz-bin/Cat?I/254"/>
      <DATA><TABLEDATA>
        <TR><TD>4.7766</TD><TD>72.8474</TD><TD>8.59</TD></TR>
        <TR><TD>5.4576</TD><TD>72.6528</TD><TD>12.18</TD></TR>
        <TR><TD>3.9867</TD><TD>72.9484</TD><TD>12.09</TD></TR>
        <TR><TD>8.9587</TD><TD>72.6635</TD><TD>14.38</TD></TR>
        <TR><TD>5.4847</TD><TD>72.8272</TD><TD>14.96</TD></TR>
      </TABLEDATA></DATA>
    </TABLE>
  </RESOURCE>
</VOTABLE>
```

图 2.1: VOTable 格式文档实例

## 2.2.4 VOTable 模式

图 2.2 所示是 VOTable 1.0 版的模式结构。

## 2.2.5 目前支持VOTable数据格式的解析工具和应用软件<sup>[10]</sup>

名称	实现语言	开发单位	功能描述	链接
JAVOT	Java	NVO	DOM 方式，首个发布解析器， 仅支持读，解析速度一般	<a href="http://www.us-vo.org/VOTable/JAVOT/">http://www.us-vo.org/VOTable/JAVOT/</a>
SAVOT	Java	European VO	SAX 方式，支持 VOTable 的 读、写、编辑，解析效率最好	<a href="http://simbad.u-strasbg.fr/public/cdsjava.gml">http://simbad.u-strasbg.fr/public/cdsjava.gml</a>

VOTable Java Streaming Writer	Java	VO-India	支持写, 采用流的方式写文件, 内存占用少, 功能有缺陷	<a href="http://vo.iucaa.ernet.in/~voi/votableStreamWriter.htm">http://vo.iucaa.ernet.in/~voi/votableStreamWriter.htm</a>
VOTable Java Parser based on XML Schema	Java	NVO	支持 VOTable 读写	<a href="http://spider.ipac.caltech.edu/staff/jchavez/public/votable_parser.html">http://spider.ipac.caltech.edu/staff/jchavez/public/votable_parser.html</a>
Starlink Tables Infrastructure Library	Java	Starlink	支持 VOTable 读写	<a href="http://andromeda.star.bris.ac.uk/~mbt/stil/">http://andromeda.star.bris.ac.uk/~mbt/stil/</a>
C++ Parser	C++	VO-India	支持 VOTable 读	<a href="http://vo.iucaa.ernet.in/~voi/cplusplusparser.htm">http://vo.iucaa.ernet.in/~voi/cplusplusparser.htm</a>
VOTable Perl Modules	Perl	NVO	支持 VOTable 读写	<a href="http://heasarc.gsfc.nasa.gov/classx/pub/votable/">http://heasarc.gsfc.nasa.gov/classx/pub/votable/</a>
VOTable::DOM	Perl	NVO	支持 VOTable 读写	<a href="http://monet.ncsa.uiuc.edu/~rplante/VO/VOTable-DOM.pm">http://monet.ncsa.uiuc.edu/~rplante/VO/VOTable-DOM.pm</a>

表 2.4: VOTable 解析器

名称	实现语言	开发单位	功能描述	链接
VOPlot	Java	VO-India	VOTable 二维可视化工具	<a href="http://vo.iucaa.ernet.in/~voi/voplot.htm">http://vo.iucaa.ernet.in/~voi/voplot.htm</a>
Mirage	Java	Bell Labs	VOTable 可视化和分析	<a href="http://www.bell-labs.com/project/mirage/">http://www.bell-labs.com/project/mirage/</a>
Topcat	Java	Starlink	VOTable 可视化和编辑	<a href="http://www.starlink.ac.uk/topcat/">http://www.starlink.ac.uk/topcat/</a>
TreeView	Java	Starlink	天文数据结构可视化工具, 强大的集成工具	<a href="http://www.star.bristol.ac.uk/~mbt/treeview/">http://www.star.bristol.ac.uk/~mbt/treeview/</a>
conVOT	Java	VO-India	将 ASCII 文本格式和 FITS 格式文件转换成 VOTable 的工具	<a href="http://vo.iucaa.ernet.in/~voi/conVOT.htm">http://vo.iucaa.ernet.in/~voi/conVOT.htm</a>
VOTFilter	Java	China-VO	将 VOTable 导入导出 OpenOffice 的插件	<a href="http://services.china-vo.org/vofilter/">http://services.china-vo.org/vofilter/</a>

表 2.5: 支持 VOTable 的应用软件

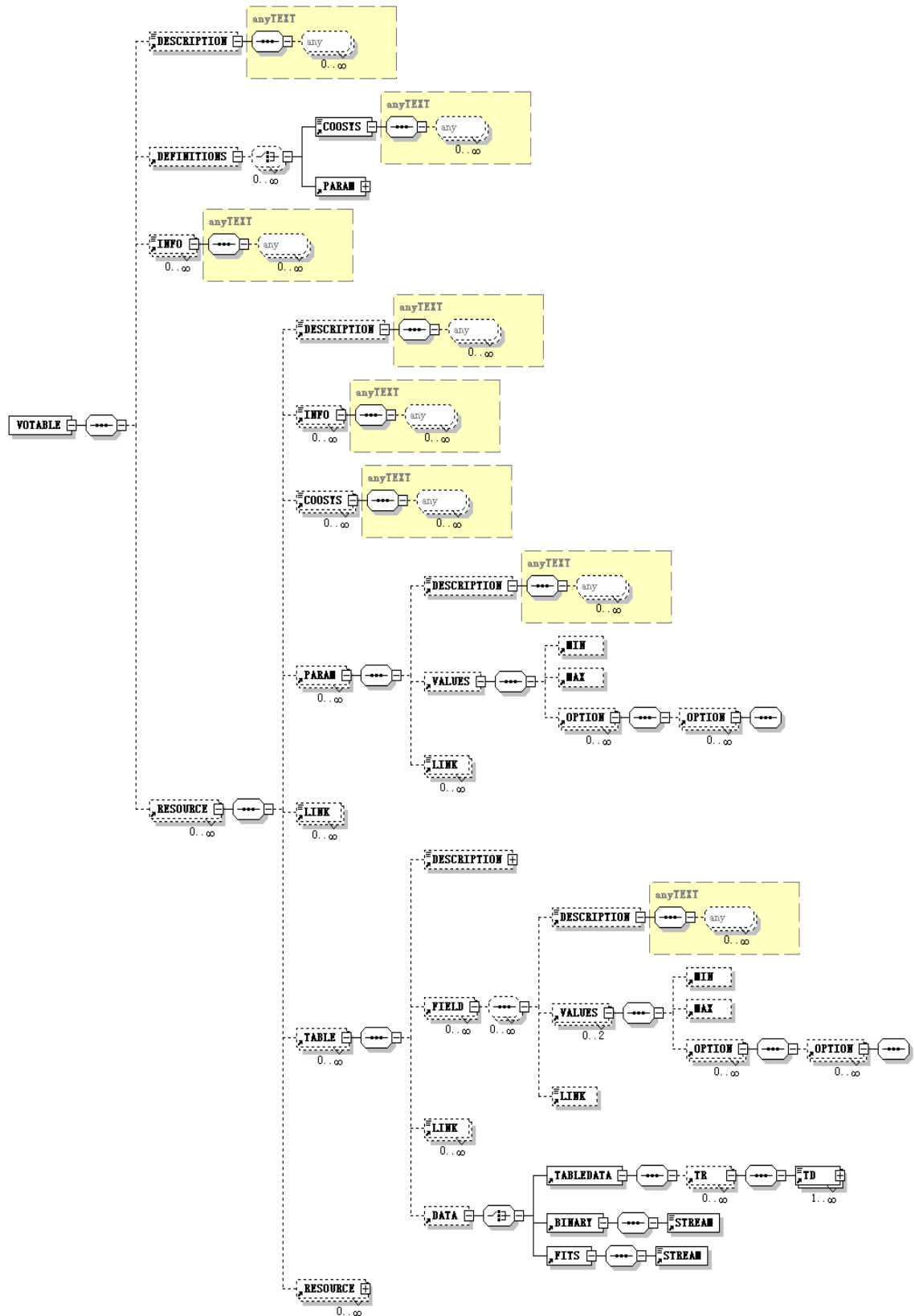


图 2.2: VOTable 1.0 Schema 图表

## §2.3 统一内容描述符 (UCD)

### 2.3.1 UCD 概述

UCD<sup>[11]</sup>(Unified Content Descriptor)是用来描述天文数据含义的一组正式的词汇表,这个词表的主要目的是使得不同数据集间的互操作更加容易,它为天文带入了一些语义网的概念,这一组词汇表是对数据量(比如星表中的列)进行用定义的词汇进行统一的描述,即描述这个物理量表示哪一类概念意义,而不是定义这个物理量的名字和单位。这一组词汇是受控制的,目的是避免出现含糊的或类似意义的UCD,并且是灵活的,它能够被人和计算机都能理解其表达数据量的意义,重要的是让计算机理解语义来实现天文的某些“人工智能”<sup>[12,13,14,15,16,17,18]</sup>。

UCD最早是由CDS与ESO合作创建,后来被称为UCD1,目的是为了统一描述星表数据列的物理意义,被应用在CDS Vizier数据库中描述其收集的星表的数据列。随后被应用到VOTable标准中,以及简单图象访问协议SIAP, Cone-Search等早期虚拟天文台颁布的标准和一些比如Aladin等应用工具中,但后来其显露出设计的一些漏洞和缺陷,UCD目前正在经历一场变革,IVOA目前正致力制订UCD 3标准,计划对UCD进行重新结构化。近期还推出了UCD1+版本,对UCD1进行修改,并已经在CDS VizierR成功测试,而且还推出了从UCD1词汇表到UCD1+的转换关系。现在由于UCD1+和UCD3还未进入广泛实用阶段,目前我们工作研究的基础还是UCD1,所以对其进行详细的介绍,而对工作草案UCD1+和UCD3做简单说明。

### 2.3.2 UCD1 的起源和应用<sup>[12,13]</sup>:

CDS(斯特拉斯堡数据中心)目前负责着Vizier服务,它收集了超过3000个星表。Vizier服务的主要组成部分是处理对这些星表进行查询的关系数据库。一个星表由一个或多个表组成,在Vizier中的这些表总共包括超过 $10^5$ 个属性列。因为这些天文星表有着不同的来源,时间跨度很大,对于这些星表的描述各种各样。表的每一个列由一个标志(列名)和单位描述这个物理量是如何被测量或如何被导出的。但

是这些描述并不精确和一致,比如恒星在V波段的星等可能在一个星表中用列名“V”表示,而同样意义的物理量却在另一个星表中用列名“Vmag”表示。我们在VizieR找到了120个不同的列名来表示这个物理量。如何才能自动地理解处理这些星表的内容,需要对这些星表结构内容进行统一的描述(也就是统一化的元数据)

为了精确描述星表的这些列的内容意义,CDS与ESO合作创建了UCD,后来被称为UCD1。他们采取了从底向上的方法大约定义了1394条词汇(UCD)来描述 $10^5$ 个列。这些UCD词汇以层次结构排列,共包括4层,尝试根据物理量的意义分成几大类。在结果树中,主要的几大分支包括:PHOT(天体光度相关的参数),POS(天体位置相关的参数),INST(观测设备参数)。

UCD词汇的层次性结构并不是强制的,它仅仅是为了更好的描述整个UCD词汇集。重要的一点是UCD提供了天文中大部分能被测量和处理的的基本语义描述,因为它是来源于大量的数据描述,采用从底向上的方法制定的。

UCD的语法是:每一个UCD都是一个字符串,一般包含2-4个层次,每一个层次由一个大写的单词表示,不同的层次由下划线分隔。比如一个星表中唯一的表示赤经的列的UCD就是POS\_EQ\_RA\_MAIN。

UCD是一个非常有用的元数据用于星表的挑选,因为UCD提供了一个基于星表内容进行精确选择的机制。如若没有UCD,想要查找K波段星等(2.2  $\mu\text{m}$  波长附近的亮度),他或许得花费时间把红外和近红外的所有星表的描述看一遍,如果星表的描述不太精确的话,他甚至会遗漏掉有用的星表。使用UCD,可以很容易的导出所有包含K波段星等测量的星表列表,这仅仅靠查找包括UCD=“PHOT\_MAG\_K”或UCD=“PHOT\_JHN\_K”列的星表。

UCD已经被用在了VOTable文档中,用来描述列信息的<FIELD>元素包含ucd属性,以及列名,单位和数据类型等属性。ucd属性即这一列的UCD,提供了对数据列标准简练的语义内容描述。

UCD被认为最有前景的应用是自动对照两个不同表的内容。如果两个表的UCD都可以得到。那么很容易确定那些列包含着同样的物理量。我们已经开发了一个原型基于UCD对两个VOTable文档进行连接查询(join)。如果两个文档包括相同UCD的列,那么就会建议对这两个列进行比较,如果需要,还自动进行单位的转换。这在星表的交叉认证中将会得到巨大应用。

UCD 还用于星表的可视化显示上，如虚拟天图的显示，我们开发了一个程序自动的找出星表内表示位置信息的列，以及星等，根据这些生成一张天图。

UCD 由于是处于开始阶段，不能涵盖所有的天文语义是肯定的，比如没有包括图象参数和数据压缩的描述。而且现在认为最大的缺陷在于将所有的误差的 UCD 都定义为 **ERROR**，没有对误差的物理意义做任何描述，这带来了很大的麻烦。目前对于我们来说，UCD1 的出现使一些应用成为可能，但 UCD1 的缺陷又使得我们在应用的实现上遇到了一些障碍。

### 2.3.3 UCD3:

由于 UCD 的巨大意义和应用前景，以及目前 UCD1 的缺陷，IVOA 成立了专门的 UCD 工作组来对 UCD 标准进行制订和修改，目前计划制订 UCD3 标准，并把它作为实现将来天文语义网格努力的一部分。UCD3 在结构上重新构造 UCD 层次树，语意更加完整，具体见。并且在语法上也做了很大变动，舍弃原来大写单词加下划线的语法构造方式（如 `POS_EQ_DEC_MAIN`），而是由多个单词中间用分号隔开组成一个 UCD 字串，而单词由用句点隔开的单字（atom）组成（如 `meta.id;src` 或 `arith.ratio;phot.flux;em.radio`）。标准的经过 IVOA 认可的 UCD 可以由名称空间 `ivoa:` 作为前缀（如 `ivoa:arith.ratio;ivoa:phot.flux;ivoa:em.radio`），但名称空间 `ivoa:` 是可选的，可以省略。有关 UCD3 的详细资料参见[15,16,17,18]。

### 2.3.4 UCD1+:

目前UCD3 还在研究制订之中，目前的进展比较缓慢，短时间内不可能投入使用，为了满足目前应用的需要，IVOA部分参与者又提出了新的UCD1+建议，即对 UCD1 进行修订，其词汇构成语法与UCD3 一致，兼容未来的UCD3<sup>[19,21]</sup>，并且能很快投入使用来替代UCD1，目前将UCD1 原有的 1394 词汇个映射成 602 个UCD1+词汇（映射表参见[22]），新的UCD1+更加灵活，表达的语义更加精确，比如将UCD1 的误差描述（**ERROR**）细化，设计了一系列表示误差意义的UCD1+新词汇。目前 UCD1+已经在 VizierR 上进行了初步测试，表明UCD1+具有较大的应用前景，我们也在中国虚拟天文台的具体应用中也及时引入了UCD1+。

## § 2.4 天文数据查询语言 ADQL

### 2.4.1 ADQL 概述

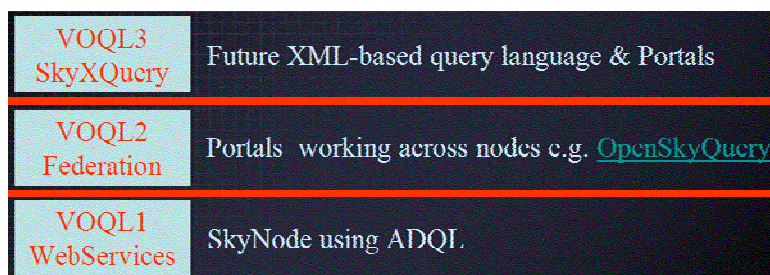
ADQL<sup>[23]</sup>(Astronomical Data Query Language)被设计用来描述查询表列结构天文数据要求的查询语言<sup>[24]</sup>，它基于结构化查询语言(SQL)，采用XML文档格式，它实际上是一种XML格式的SQL语言。ADQL设计的重要目的是作为天文统一的数据查询语言，实现天文数据的分布式查询，目前 90%天文数据库是关系型数据库，而且大部分天文表列格式的数据是由各种关系型数据库管理的，ADQL最重要的是屏蔽了各关系数据库平台实现SQL语言标准的不同（许多数据库管理系统对SQL语言做了不同的扩展），它可以对存储在各种数据库中的天文数据的查询进行统一的描述，为实现多数据库平台下的天文数据统一查询访问和互操作提供了解决途径。而且为了弥补SQL缺乏描述空间区域以及时空坐标等天文专用的语句，满足方便简洁描述区域查询和坐标变换的要求，ADQL还提供了一些描述区域查询、交叉证认和时空坐标系的语法，并且计划在将来做更大的扩展和完善，以满足查询表列数据以外的复杂数据如图象数据和光谱数据的要求，比如引入UCD，增加新的数据类型和操作等等。目前ADQL已经成功地应用到SkyQuery中和很多项目中，我们也采用ADQL作为星表查询服务的输入查询语言。

<pre> &lt;?xml version="1.0" encoding="utf-8" ?&gt; - &lt;Select xmlns:xsd="http://www.w3.org/2001/XMLSchema"   xmlns:xsi="http://www.w3.org/2001/XMLSchema-   instance" xmlns="http://adql.ivoa.net/v0.73"&gt; - &lt;SelectionList&gt;   &lt;Column Table="t" Name="b" /&gt;   &lt;Column Table="b" Name="*" /&gt; &lt;/SelectionList&gt; - &lt;From&gt;   &lt;Table Name="Tab" Alias="t" /&gt;   &lt;Table Name="Bob" Alias="b" /&gt; &lt;/From&gt; - &lt;Where&gt; - &lt;Compare Comparison="&lt;"&gt;   - &lt;Function&gt;     &lt;Aggregate&gt;AVG&lt;/Aggregate&gt;     &lt;Column Table="b" Name="dd" /&gt;   &lt;/Function&gt;   - &lt;Atom&gt;     &lt;Real&gt;4.56&lt;/Real&gt;   &lt;/Atom&gt; &lt;/Compare&gt; &lt;/Where&gt; - &lt;OrderBy&gt; - &lt;Item&gt;   &lt;Column Table="b" Name="dd" /&gt; &lt;/Item&gt; &lt;/OrderBy&gt; &lt;/Select&gt; </pre>	<p>==</p>	<pre> <b>Select t.b,b.* from Tab t, Bob b Where Avg(b.dd) &lt; 4.56 order by b.dd</b> </pre>
---	-----------	--

表 2.6: ADQL的一个例子以及对应的SQL语句<sup>[25]</sup>

### 2.4.2 ADQL 的发展和未来

ADQL 的格式和用途现在仍处在激烈的讨论中，目前还没有正式版本推出，目前最新的是 0.7.4 版本，它仅实现了 SQL 中的 Select 语句，在将要推出的 0.8 版本中实现了 SQL 中数据操纵语言(DML)中 Insert, Delete, Update 语句以及支持对用户定义函数、SQL 注释和子查询的描述。IVOA 计划将 ADQL 作为 VOQL ( Virtual Observatory Query Language)语言的最底层，计划中的 VOQL 由三层组成，其目的是作为最终的虚拟天文台的标准查询语言，VOQL 结构如下：

图 2.3: VOQL的三层体系结构<sup>[26]</sup>



## § 2.5 虚拟天文台资源元数据标准 (Resource Metadata 1.0)

### 2.5.1 虚拟天文台资源元数据标准<sup>[27]</sup>概述:

虚拟天文台的一个重要的功能是提供可获取的数据资源和计算资源的定位, 并且告诉用户如何调用它们, 我们需要收集数据集和数据服务的元数据, 这样虚拟天文台用户能方便地找到他们感兴趣的数据和信息。另外, 高效管理分布式的查询也需要这些元数据, 如果用户对X射线图象感兴趣, 那么就没有必要去查询HST的图象数据库<sup>[28,29]</sup>。

现在 IVOA 已于 2004 年 1 月 24 日推出资源元数据标准正式版本 1.0 版, 它给出了资源元数据和服务元数据的概念和基本框架, 以及如何具体应用。资源元数据给出了一组对资源基本属性和内容描述的关键词, 其中有部分是必须的, 一部分是可选的。服务元数据是基于 Web 的, 包括服务接口元数据和服务功能元数据。服务接口元数据描述如何输入和输出, 也就是如何获取服务, 服务功能元数据描述服务提供的功能和性能限制。

资源 (Resource) 在 VO 中是一个非常宽泛的词汇, 几乎所有东西都可称为资源。一个资源是一个或多个服务或其它资源的集合。资源有一些共同的元数据特征, 比如发布者、创立者、标识、类型等。资源 (Resource) 在 VO 中是一个非常宽泛的词汇, 几乎所有东西都可称为资源。一个资源是一个或多个服务或其它资源的集合。资源有一些共同的元数据特征, 比如发布者、创立者、标识、类型等<sup>[30]</sup>。

服务的定义如下: 一个服务是任何能被用户访问并根据其自身能力执行某种活动的资源。服务的元数据包括用户需要决定一个服务是否是他感兴趣的信息, 以及告诉用户如何调用这个服务。

为了使天文信息服务能够融入虚拟天文台, IVOA 推出了一个资源元数据管理的多层次系统, 在顶层, 我们需要尽可能少的信息, 仅仅能说明某个资源的存在和谁负责管理这个资源就足够了, 在底层, 资源元数据将会更广泛和复杂。

### 2.5.2 资源元数据概念

所有的资源通过元数据来描述，资源元数据是独立于任何专门的服务的。资源元数据包括如下几部分：

1. 标识元数据 (Identity metadata)：包含最少的信息，给出资源的标题，缩写名和一个唯一的标识，条目如下：

中文名称	英文名称	类型	解释说明
标题	Title	string	资源名称
缩写名	ShortName	String	资源名称缩写
标识	Identifier	URI	一个唯一的标识 URI

2. 履历元数据(Curation metadata)：描述谁支持这个资源以及可获取性，比如版本，发布时间，发布者，创建者和贡献者等，具体条目如下：

中文名称	英文名称	类型	解释说明
发布者	Publisher	string	发布资源的实体，对资源的可用性负责
发布者标识	PublisherID	URI	资源发布者的标识，可以是一个网络链接
创建者	Creator	string	资源主要内容的创建实体，可以是个人或组织
创建者标识	Creator.Logo	URL	指向标识的 URL，用于标识信息资源
贡献者	Contributor	string	对资源内容有贡献的实体
日期	Date	string	在资源的生命期中的一个重要日期，比如创建或发布日期

版本	Version	string	与资源创建或者发布相关的一个标号
联系方式	Contact		资源负责人，包括联系人姓名或电子邮件地址
联系名称	Contact.Name	string	联系方式的名称
联系电邮	Contact.Email	e-mail	联系方式的电子邮件

### 3. 内容元数据(General content metadata): 概要描述资源的内容。

中文名称	英文名称	类型	解释说明
主题	Subject	string, list	对资源的主题、天体类型或其他属性进行描述的一系列关键词
来源	Source	string	资源来源，从何处导出的本资源
资源参考	ReferenceURL	URL	一个指向资源额外信息的 URL 地址
类型	Type	string, list	资源内容的种类或类型
内容等级	ContentLevel	string, list	内容等级描述，指出资源的潜在用户
联系	Relationship	string	和其他资源相关的资源
联系标识	RelationshipID	URI	相关资源标识

### 4. 收集和服务内容元数据(Collection and service content metadata): 描述这个资源数据集的专门属性。

中文名称	英文名称	类型	解释说明
机构	Facility	string, list	获取数据的天文台或研究机构
设备	Instrument	string, list	收集数据所用设备
覆盖	Coverage		资源内容的覆盖范围
空间覆盖	Coverage.Spatial	string	资源的覆盖的天区
	Coverage.RegionOfRegard	float,	
覆盖波段	Coverage.Spectral	string, list	覆盖波段, 列表, 可取值: Radio, Millimeter, Infrared, Optical, Ultraviolet, X-ray, Gamma-ray
覆盖具体波段	Coverage.Spectral.Bandpass	string, list	覆盖具体波段, 列表, 例: U, B, V, R, I
观测波段中心波长	Coverage.Spectral.CentralWavelength	float	观测波段中心波长, 单位: 米
观测波段最小波长	Coverage.Spectral.MinimumWavelength	float	观测波段最小波长, 单位: 米
观测波段最大波长	Coverage.Spectral.MaximumWavelength	float	观测波段最大波长, 单位: 米
起始时间	Coverage.Temporal.StartTime	string	观测开始时间, 一般精确到日期
终止时间	Coverage.Temporal.StopTime	string	观测结束时间, 一般精确到日期
覆盖极限灵敏度	Coverage.Depth	float	观测深度
天体密度	Coverage.ObjectDensity	float	资源的目标密度, 以每平方度计数
天体计数	Coverage.ObjectCount	int	资源包括的目标数目
覆盖天区比例	Coverage.SkyFraction	float	覆盖天区的比例, (0, 1]
分辨率	Resolution		

空间分辨率	Resolution.Spatial	float	空间(角)分辨率, 单位: 十进制度
光谱分辨率	Resolution.Spectral	float	光谱分辨率, $\Delta \lambda / \lambda$ 比 率值
时间分辨率	Resolution.Temporal	float	时间分辨率, 单位: 秒
UCD	UCD	string, list	有关的 UCD 列表
格式	Format	string, list	资源所提供信息的物理 或数字表示方式
权限	Rights	string	资源的所属及访问权限

### 2.5.3 服务元数据概念(Service metadata)

服务元数据包括描述服务接口的元数据和服务能力的元数据, 服务接口元数据描述如何获取这个服务, 即如何输入和输出。这个接口有可能是基于 Web 浏览器的接口(如 HTML 表格), Web 服务接口(采用 WSDL 描述), 或 Grid Service 接口, 以及 GLU 描述接口等等。服务能力元数据包括描述服务提供功能, 限制以及其他行为特征。具体说明如下:

#### 1. 服务接口元数据(Interface metadata):

中文名称	英文名称	类型	解释说明
服务接口 URL	Service.InterfaceURL	URL	一个指向服务接口说明文当 的 URL
服务基准 URL	Service.BaseURL	URL	用户激活服务的 URL 的基本 部分
HTTP 服务接口返回结果	Service.HTTPResultsMIMEType	MIME type	服务返回结果的 MIME 类型

## 2. 服务能力元数据(Capabilities metadata):

---

中文名称	英文名称	类型	解释说明
服务标准 URI	Service.StandardURI	URI	标识一个服务标准的 URI
服务标准 URL	Service.StandardURL	URL	一个指向服务标准描述的 URL
服务最大搜索半径	Service.MaxSearchRadius	float,	服务提供者或服务附加的最大搜索范围，以度的形式规定最大的搜索半径
服务最大返回记录个数	Service.MaxReturnRecords	int	

---

## 参考文献

- [1] Open Discussion Lists of IVOA Working Groups. <http://www.ivoa.net/forum/>
- [2] François Ochsenbein, Roy Williams. VOTable: A Proposed XML Format for Astronomical Tables. <http://cdsweb.u-strasbg.fr/doc/VOTable/votable-1-0.htx>
- [3] [Astrores] Vizier. Describing Astronomical Catalogues and Query Results with XML. <http://vizier.u-strasbg.fr/doc/astrores.htx>
- [4] [XSLT] XSL Transformations. <http://www.w3.org/TR/xslt>
- [5] François , Roy Williams. VOTable Format Definition Version 1.05. <http://www.ivoa.net/internal/IVOA/IvoaVOTable/VOTable-1-1.pdf>
- [6] François, Roy Williams. VOTable Format Definition Version 1.093 <http://www.ivoa.net/internal/IVOA/IvoaVOTable/VOTable-1.093.pdf>
- [7] François, Roy Williams. VOTable Format Definition Version 1.091 <http://www.ivoa.net/internal/IVOA/IvoaVOTable/VOTable-1.091.pdf>
- [8] François, Roy Williams. VOTable Format Definition Version 1.09 <http://www.ivoa.net/internal/IVOA/IvoaVOTable/VOTable-1.09.pdf>
- [9] François, Roy Williams. VOTable Format Definition Version 1.094 <http://www.ivoa.net/internal/IVOA/IvoaVOTable/VOTable-1.094.pdf>
- [10] VOTable Software . <http://www.ivoa.net/twiki/bin/view/IVOA/VOTableSoftware>
- [11] [UCD] Unified Content Descriptor. <http://vizier.u-strasbg.fr/UCD/>
- [12] S. Derriere. U C D : Status and perspectives. <http://www.ivoa.net/internal/IVOA/IvoaUCD/UCD-status0203.pdf>
- [13] S. Derriere. UCDs: metadata for astronomy. [http://www.ivoa.net/internal/IVOA/IvoaUCD/derriere\\_ESSW03.pdf](http://www.ivoa.net/internal/IVOA/IvoaUCD/derriere_ESSW03.pdf)
- [14] Roy Williams. Report of UCD Working Group. <http://www.ivoa.net/internal/IVOA/IvoaUCD/Cambridge-ucd.pdf>
- [15] Sebastien Derriere, Norman Gray. UCD (Unified Content Descriptor). <http://www.ivoa.net/internal/IVOA/IvoaUCD/UCD-1.9.3.pdf>
- [16] Sebastien Derriere, Norman Gray. UCD (Unified Content Descriptor). <http://www.ivoa.net/internal/IVOA/IvoaUCD/UCD-1.9.4.pdf>

- [17] Sebastien Derriere, Norman Gray. UCD (Unified Content Descriptor).  
<http://www.ivoa.net/internal/IVOA/IvoaUCD/UCD-1.9.9.pdf>
- [18] Sebastien Derriere, Norman Gray. Metadata Content within VO Resources.  
<http://www.ivoa.net/internal/IVOA/IvoaUCD/UCD-1.9.9b.pdf>
- [19] Sebastien Derriere, Norman Gray. UCD (Unified Content Descriptor) - moving to UCD1+ Version 1.03.  
<http://www.ivoa.net/internal/IVOA/IvoaUCD/WD-UCD-20040426.pdf>
- [20] Sebastien Derriere, Andrea Preite Martinez. Electromagnetic spectrum description in UCD.  
<http://www.ivoa.net/internal/IVOA/IvoaUCD/NoteEMSSpectrum-20040520.pdf>
- [21] Andrea Preite Martinez, Sebastien Derriere. The UCD1+ controlled vocabulary.  
<http://www.ivoa.net/internal/IVOA/IvoaUCD/WD-UCDlist-20040520.pdf>
- [22] Reference IVOA UCD1+ Lists. <http://vizier.u-strasbg.fr/UCD/lists/>
- [23] [ADQL] <http://www.ivoa.net/internal/IVOA/IvoaVOQL/ADQL-0.7.4.xsd>
- [24] IVOA VOQL Working group. IVOA Astronomical Data Query Language Version 0.7.4. <http://www.ivoa.net/internal/IVOA/IvoaVOQL/ADQL-0.7.4.pdf>
- [25] [http://www.ivoa.net/internal/IVOA/IvoaVOQL/adql\\_v073\\_examples.zip](http://www.ivoa.net/internal/IVOA/IvoaVOQL/adql_v073_examples.zip)
- [26] <http://www.ivoa.net/twiki/bin/view/IVOA/IvoaVOQL>
- [27] Resource Metadata for the Virtual Observatory.  
<http://www.ivoa.net/Documents/PR/ResMetadata/ResMetadata.html>
- [28] Resource Metadata for the Virtual Observatory Version 1.0.  
<http://www.ivoa.net/Documents/PR/ResMetadata/ResMetadata.html>
- [29] Raymond L. Plante. A Scalable Metadata Framework for the Virtual Observatory.  
<http://us-vo.org/pubs/files/fw-draft2.pdf>
- [30] 崔辰州. 博士学位论文 中国虚拟天文台系统设计. 2003:94-97





## 第三章 天文星表数据库的设计

### § 3.1 天文数据中心的发展

#### 3.1.1 天文数据中心概述

国家天文台目前承担着世界数据中心天文中心也称北京天文数据中心 (BADC<sup>[1]</sup>) 的建设和维护工作。世界数据中心天文中心从八十年代开始进行全国性的天文数据服务工作, 九十年代后随着计算机和网络技术的发展, 天文数据服务环境有了很大的改善。目前已建立了自己的天文数据库和网络数据服务系统, 并通过国际交换和网络下载积累了大量的适合我国天文研究需要的天文数据。1996 年底开始建立了一套天文数据的网络服务系统, 包括一个匿名 FTP 文件传输服务和 WWW 网络数据信息服务系统。天文数据的服务范围不断扩大, 其用户包括北京、上海、南京、云南、陕西、新疆等天文台站和全国各大院校的天文工作者, 以及地球和空间等学科的研究人员, 获得了很好的效果, 比如在太阳活动预报、卫星轨道预报以及大地测量与归算等。

BADC 是世界数据中心 (WDC<sup>[2]</sup>) 中唯一的天文学科数据中心, 在国际上与美国、法国、俄罗斯和日本的天文数据中心并列, 并与这些中心有相互交换和共享数据的协议。BADC 可以免费得到国际上的各种天文数据, 而且拥有大量自产数据 (如国内各个望远镜的观测数据) 和各种文献资料以及大科学工程 (如 LAMOST 项目) 投入使用后将产生的海量观测数据。

天文数据中心目前起着非常重要的作用, 集中体现在以下三个方面:

(1) 作为中国目前唯一的天文数据中心, 为国内天文界和其他用户提供天文数据服务。目前天文数据中心提供天文数据的查询和下载等一般性的服务, 还提供着多波段星表交叉证认等复杂的服务。

(2) 天文数据中心积累的天文数据库是虚拟天文台的基础和研究开发的平台和实验床。天文数据中心积累的天文数据和建设的天文数据库是虚拟天文台的根基, 天文数据的丰富和天文数据库的完善直接影响虚拟天文台的功能和研发难度。

(3) 为 LAMOST 输入星表的制定提供丰富的数据源。LAMOST 的观测目标的

预选是一个非常重要的工作，它直接决定 LAMOST 项目的科学产出。只有基于最丰富的数据资源，才能选出最有价值的观测目标，制定出最精确的目标星表和导星星表。

### 3.1.2 天文数据中心的发展现状

经过多年的建设，特别是 2001 年获得资金支持后，天文数据中心的数据的收集整理量、软硬件系统和网站建设上都取得了飞跃式的发展。

#### 天文数据中心的数据资源

下图显示的是从 2002 年 5 月到 2004 年 5 月的数据增长情况，天文数据中心的数据量已经从 2001 年初的 470GB 增长到 2004 年 5 月的 2400GB，3 年增长了 5.1 倍。

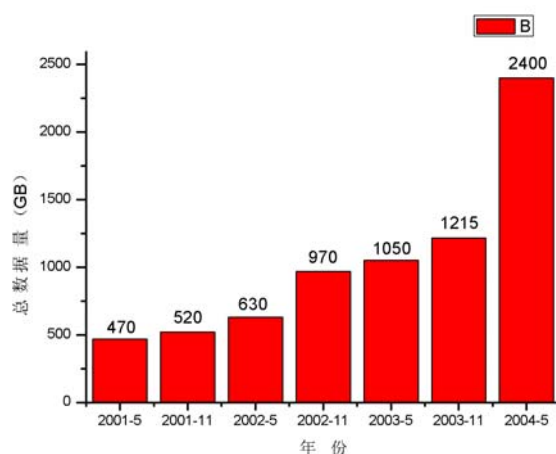


图 3.1: BADC 天文数据的增长

天文数据的种类和质量也有大幅度的提高，我们特别重视国内外重要的高质量天文数据的收集整理，我们设法在最短的时间内将国外公布的高质量数据收集到手，然后对其进行二次处理和归档，以最快的速度向国内天文学家发布。像 2MASS 全天星表是 UMASS 在 2003 年 3 月 25 日释放，而我们在 5 月底前整理归档完毕并对外公布。我们在 2002 年 5 月和 2003 年 12 月分别对国内的 BATC 巡天数据和美国 SDSS DR1<sup>[3]</sup> 数据库做了镜像和再处理，这在国内是唯一的。下表列出的是截止到 2004 年

5 月份中国天文数据中心整理收集的天文专业数据。

类别	数据集名称	数据量（压缩）
Catalog	CDS/ADC 星表	≈30GB
	其他大星表 (2MASS GSC2.2,USNO B1.0)	160GB
Survey	RealSky 天图	5GB
	DSS-I 天图	68GB
	DSS-II 天图	≈520GB
	BATC	340GB
	SDSS EDR	30GB
	SDSS DR1	>900GB
	SDSS DR2 Spectro	62GB
	2dF 2003+2dF QSO	8GB
	Swiss Ephemeris	3GB
Archive	GOODS	40GB
	HyperCat	5GB
	ROSAT+RASS	38GB
	Einstein X 射线卫星数据	5GB
Library	ADS 文献	>350GB
	古代天象记录	1GB
总计		约 2400GB

表 3.1: 目前天文数据中心拥有的天文数据

### 天文数据中心的硬件建设

2000 年: ADS 文献镜像服务器建成, 2003 年对镜像服务器做了全面升级。

2002 年 4 月: 完成了天文数据中心数据主服务器的搭建, 实现了 4TB 的存储空间, 如图所示。我们没有直接购买 IT 厂商提供的数据服务器, 而是采用了一种

DIY 的方式，自行购买部件，自行搭建服务器，总投资仅有 28 万元，是同容量品牌服务器价格的四分之一到三分之一。这台数据服务器的主要性能指标如表 2 所示：

计算处理能力	内存	数据空间	网络通讯
Intel P4 1.2GHz×10	1GB×5	80GB+20GB×4+1TB×4	D-link Switcher
总造价	28 万		

表 3.2 天文数据中心数据服务器主要性能表

2003 年 1 月：完成了 SDSS EDR 数据服务器的搭建。

2004 年 1 月：完成了 SDSS DR1 数据服务器的搭建，替代了 SDSS EDR 数据服务器。



图 3.2：数据中心服务器

### 天文数据中心的软件系统和网站建设

天文数据中心数据主服务器采用 LINUX 操作系统。主要好处有两点：一方面，由于 LINUX 操作系统和其平台上绝大部分应用程序都实行开放源码政策，这有利

于数据中心应用程序的开发和调试；另一方面，LINUX 操作系统和其平台上绝大部分应用程序都通过网络发行，可以节省数据中心许多软件方面的开支。

SDSS DR1 数据服务器采用的是 Win2K Server 操作系统，数据库管理系统采用的是 SQL Server 2000。

目前，数据中心网站 (<http://badc.lamost.org>) 已经开始对外服务。现在网站主要的内容包括：天文数据库系统、中国虚拟天文台计划、业余天文服务器、中国天文网络与软件。从天文数据库系统可以找到数据中心已经对外服务的各类天文数据。

### 3.1.3 天文数据分类标准

根据天文数据的特点，分类时首先要考虑分类的科学性和系统性。也就是要遵循层次性原则，组成一个既有纵向从属关系，又有横向协调补充关系分类系统。还应该考虑可扩展性和兼容性，由于观测手段和观测方式的不断发展，目前的分类系统是不能穷尽所有的数据，因此在系统的每个层次上都应该具有扩展性和兼容性。最后，分类时要尽可能考虑分类系统的实用性，天文数据中心的发展方向不仅仅是简单的数据共享，而是逐步走向虚拟天文台。因此实用性在分类中占据举足轻重的地位。

为了便于天文数据信息资源的组织、管理、检索和利用，将数据分为四级。

**一级**为基本类目，用单个大写英文字母表示。

**二级**为一级数据的进一步细分。类目代码采用两位数字表示。

**三级**为二级数据的进一步细分。类目代码采用一位数字表示。**D** 类只到二级不再划分。

**四级**为三级数据的进一步细分。类目代码采用一位数字表示。

一级类目	二级类目	三级类目	四级类目
A-专业数据	A01-巡天观测数据	A011-巡天测光数据	A0111-DSS I
			A0112-USNO
			A0113-DSS II
			A0114-SDSS
			A0115-60/90cmBATC 巡天观测资料
		A012-巡天光谱数据	A0121-SDSS
			A0122-2df
			A0123-LAMOST
		A013-巡天射电数据	A0131-NVSS
			A0132-FIRST
			A0133-密云米波综合孔径射电望远镜巡天资料
		A014-巡天红外数据	A0141-2MASS
			A0142-IRAS
		A015-巡天高能数据	A0151-ROSAT
		A02-定点观测数据	A021-定点测光数据
	A0212-兴隆 2.16m 光学望远镜观测数据		
	A0213-云南 1m 光学望远镜		
	A0214-佘山 1.56m 望远镜观测数据		
	A022-定点光谱数据		A0221-兴隆 2.16m 光学望远镜观测数据
	A023-定点射电数据		A0231-德令哈观测站 13.7m 射电望远镜观测数据
	A024-定点高能数据		
	A025-定点红外数据		

	A03-太阳物理数据	A031-太阳磁场数据	A0311-怀柔太阳磁场望远镜观测数据	
		A032-太阳射电数据	A0321-10cm 太阳射电望远镜观测数据	
	A04-时变数据	A041-变源数据	A0411-脉冲星数据	
		A042-移动天体	A0421-近地天体彗星数据	
		A043-瞬态天体	A0431-超新星、 $\gamma$ 爆数据	
	A05-联合与导出数据	A051-包括距离		
		A052-化学丰度		
		A053-电磁场		
		A054-重力场		
	A06-天体位置数据	A061-天体测量数据		
		A062-交叉认证		
		A063-非恒星与展源等		
	B-应用数据	B01-日地物理数据	B011-太阳活动常规观测数据	
			B012-太阳地球物理资料	
B02-天文地球动力学		B021-地球角动量测量		
		B022-地球自转变化参数测量		
B03-轨道观测数据		B031-近地小天体与空间碎片测量		
		B032-人卫轨道观测等数据		
B04-时间历法数据		B041-行星历表	B0411-DE404	
C-文献数据		C01-中文文献	C011-天文期刊	C0111-ADS



		C012-天文书籍	
		C013-相关学科文献	
	C02-英文天文文献	C021-天文期刊	
		C022-天文书籍	
		C023-相关学科文献	
		C024-古代文献	C0241-古代天象记录
	D-天文信息	D01-教育资源	
D02-网络资源			

表 3.3 天文数据的分类标准

## § 3.2 天文星表数据库的设计

### 3.2.1 天文星表数据库概述

星表是包含天体信息的数据表格，是天文学家最常用到的天文数据，目前可以获取到的星表数目很大，而且星表的大小差别悬殊，从包括几十个目标的小星表到包含上亿甚至 10 亿目标的大星表，而且近年来巡天得到的大星表越来越多，为实现对这些星表的高效管理和查询，我们设计实现了一个功能完善的天文星表数据库管理系统。

我们设计实现的天文星表数据库的主要用途如下：

1. 作为北京天文数据中心数据库的重要部分，收集和整理重要的和最新的星表数据集，为国内天文学家和其他用户提供星表查询服务；
2. 为设计实现 LAMOST 光谱数据库积累建库经验；
3. 作为虚拟天文台的重要基础设施，为虚拟天文台提供开发基础和实验平台，是虚拟天文台数据服务的基础。我们基于星表数据搭建了虚拟天文台网格环境下第一个数据服务，并继续基于这个数据库进行服务开发和测试。并探讨未来适合虚拟天文台的数据库管理系统的结构框架和实现功能。
4. 星表数据库的一个重要的作用是进行多波段星表的交叉认证工作，目的是为

LAMOST 遴选观测目标，制定输入星表，交叉证认的另一个目的是对多波段星表进行数据挖掘研究工作。目前国际上数据库服务仅提供小数据量数据的交叉证认，像大型星表的全部交叉证认只能由我们自己来设计数据库实现。

### 3.2.2 天文星表数据库的结构组成和环境

天文星表数据库包括“数据”库和“元数据”库两部分，“数据”库是由导入的星表数据组成，也称星表数据库。“元数据”库管理收集这些星表的有关元数据信息，也称星表元数据库。这两个数据库之间是数据与元数据的关系。

星表数据库的各个表结构是由自动入库程序按照规定的入库规则形成的，其内容不断充实更新，而星表元数据库结构是采用开发工具预先设计好的，结构不再做大的变动。两个数据库部分的设计实现方式不同，所以我们分别阐述星表数据库和星表元数据库的设计实现。

天文星表数据库的设计、运行环境如下：

#### 1. 软件系统

(1) 操作系统：RedHat Linux 9.0 版

(2) 数据库管理系统：MySQL<sup>[4]</sup> 4.1 版

MySQL是一个客户机/服务器体系结构的分布式关系型数据库管理系统，它是目前可以得到的最快的数据库之一，MySQL可以应用于各种网络环境中，无论是LAN，WAN或Internet，用户都可以通过网络访问MySQL数据库，而通过Internet则可以与任何位置的任何人共享数据，MySQL数据库的可移植性很好，MySQL可以运行在各种常见的操作系统中，包括Unix，Solaris，FreeBSD，Linux，Windows，OS/2 等，所以可以方便地将数据库从一个操作系统转移到另外一个操作系统，而数据和应用不需要做修改。

MySQL是目前最受欢迎的开放源代码数据库软件，符合GNU规则，可以对数据库不适合具体应用的地方进行修改。MySQL是一个多线程、多用户的数据库产品，提供了多种编程语言如Java，C，C++，PHP，Perl的应用编程接口。但是MySQL4.1 数据库软件并未实现关系数据库的所有功能，比如R-Tree索引，外键，视图，事务处理等。MySQL是改进最快的数据库系统

之一，一直在进行着版本的更新和功能的扩展<sup>[5,6,7]</sup>。

(3) 数据库设计开发工具：采用Windows下的Sybase公司的数据库建模工具——PowerDesigner<sup>[8]</sup> 10.0 评估版，它提供了对基于MySQL 4 的数据库的设计。我们采用它设计生成了星表元数据库。

## 2. 硬件系统：

为天文数据中心主机中的一台子机 Barm2, 1.4GHz P3 双 CPU, 1G SDRAM 内存, Raid5 存储系统, 8 块硬盘, 硬盘规格 Maxtor ATA/133 160GB, Raid 总数据容量  $7*160=1120\text{GB}$ 。

### 3.2.3 星表数据来源和格式

我们入库的主要数据来源于斯特拉斯堡天文数据中心(CDS<sup>[9]</sup>)负责收集整理天文星表，这些星表包括恒星、星系，以及其他银河系外星系天体，CDS也收录太阳系内天体数据和物理数据。目前总共已经收录 5499 个星表，其中 4486 个星表通过ASCII码文本格式或FITS格式文件发布，这些文件可以通过FTP下载<sup>[10]</sup>。

#### CDS 星表文件目录结构

在 CDS 星表 FTP 地址下，每一个星表单独存储在两层目录中，最上层目录名称表示星表的类型，如下表所示，第二层目录名为序号，每一个星表都单独存储在单个目录下，例如 Tycho2 星表是天体测量星表，它的文件被存放在目录 I/259 下。目录内一般是由 ReadMe 文件+星表数据文件(文本格式)组成，ReadMe 中以标准格式提供星表的有关信息，并提供数据文件的中每一列的位置，名称单位等描述。

<b>I/number</b>	天体测量星表
<b>II/number</b>	光度测量星表
<b>III/number</b>	光谱测量星表
<b>IV/number</b>	交叉证认结果
<b>V/number</b>	合并数据
<b>VI/number</b>	混杂天体星表
<b>VII/number</b>	非星目标星表
<b>VIII/number</b>	射电星表
<b>IX/number</b>	高能星表
<b>Jlabbr/Volumn/first_page</b>	杂志上发表的星表

表 3.4: CDS 各类星表的存储目录结构<sup>[10]</sup>

### ReadMe 文件中描述数据文件结构的说明格式

ReadMe 文件中描述星表数据文件结构如下例所示，表格中的第 N 行描述的是数据文件 tycho2.dat 中第 N 列的相关信息。表格由 5 列组成，分别为 Bytes、Format、Units、Label 和 Explanation，分别描述数据文件中对应列的起止位置，数据类型、物理单位、列名称和说明。

第一列：为 Bytes，表示数据文件中对应列数据的开始和结束字节，由“-”分隔开。

如果这一列长度只有一个字节，那么只用一个数字表示开始（结束）位置，不用“-”符号。

第二列：为 Format，表示数据文件中对应列数据的格式类型，具体描述格式如下：

(a) 数据类型，用一个字母表示，有 A, I, F, E 四种（与 FITS 的数据类型相同），分别表示 Ascii 文本，整型，固定小数位数的浮点数，指数形式的浮点数。另外，极个别的星表用 D 表示浮点型数据。

(b) 紧随着的数值表示数据的宽度

(c) 最后面的点和数值表示浮点型数据的精度（仅 F, E 两种类型有）

第三列：为 Units，表示数据文件中对应列数据的物理单位，“---”表示没有单位

第四列：为 Label，表示数据文件中对应列的名称

第五列：为 **Explanation**，是对数据文件中对应列的简单说明，除说明列的具体物理意义外，有的还包含以下内容，但并不完整：

- (a) 列允许的取值范围，在方括号内列出最大最小值，
- (b) 是否这一列被排序，
- (c) 数据中空格是否被允许出现，
- (d) 一个指向更详细的解释的键（锚点）。

Byte-by-byte Description of file: <a href="#">tyc2.dat</a>					
Bytes	Format	Units	Label	Explanations	
1-	4	I4	---	TYC1	[1, 9537] += <a href="#">TYC1 from TYC or GSC (1)</a>
6-	10	I5	---	TYC2	[1, 12121] <a href="#">TYC2 from TYC or GSC (1)</a>
	12	I1	---	TYC3	[1, 3] <a href="#">TYC3 from TYC (1)</a>
	14	A1	---	pflag	[ PX] <a href="#">mean position flag (2)</a>
16-	27	F12.8	<a href="#">deg</a>	RAmdeg	[ ]? <a href="#">Mean Right Asc, ICRS, epoch=J2000 (3)</a>
29-	40	F12.8	<a href="#">deg</a>	DEmdeg	[ ]? <a href="#">Mean Decl, ICRS, at epoch=J2000 (3)</a>
42-	48	F7.1	<a href="#">mas/yr</a>	pmRA	? <a href="#">Proper motion in RA*cos(dec) (12)</a>
50-	56	F7.1	<a href="#">mas/yr</a>	pmDE	? <a href="#">Proper motion in Dec (12)</a>
58-	60	I3	<a href="#">mas</a>	e_RAmdeg	[3, 183]? <a href="#">s.e. RA*cos(dec), at mean epoch (5)</a>
62-	64	I3	<a href="#">mas</a>	e_DEmdeg	[1, 184]? <a href="#">s.e. of Dec at mean epoch (5)</a>
66-	69	F4.1	<a href="#">mas/yr</a>	e_pmRA	[0.2, 11.5]? <a href="#">s.e. prop mot in RA*cos(dec) (5)</a>
71-	74	F4.1	<a href="#">mas/yr</a>	e_pmDE	[0.2, 10.3]? <a href="#">s.e. of proper motion in Dec(5)</a>
76-	82	F7.2	<a href="#">yr</a>	EpRAm	[1915.95, 1992.53]? <a href="#">mean epoch of RA (4)</a>
84-	90	F7.2	<a href="#">yr</a>	EpDEm	[1911.94, 1992.01]? <a href="#">mean epoch of Dec (4)</a>
92-	93	I2	---	Num	[2, 36]? Number of positions used
95-	97	F3.1	---	q_RAmdeg	[0.0, 9.9]? <a href="#">Goodness of fit for mean RA (6)</a>
99-	101	F3.1	---	q_DEmdeg	[0.0, 9.9]? <a href="#">Goodness of fit for mean Dec (6)</a>
103-	105	F3.1	---	q_pmRA	[0.0, 9.9]? <a href="#">Goodness of fit for pmRA (6)</a>
107-	109	F3.1	---	q_pmDE	[0.0, 9.9]? <a href="#">Goodness of fit for pmDE (6)</a>
111-	116	F6.3	<a href="#">mag</a>	BTmag	[2.183, 16.581]? <a href="#">Tycho-2 BT magnitude (7)</a>
118-	122	F5.3	<a href="#">mag</a>	e_BTmag	[0.014, 1.977]? <a href="#">s.e. of BT (7)</a>
124-	129	F6.3	<a href="#">mag</a>	VTmag	[1.905, 15.193]? <a href="#">Tycho-2 VT magnitude (7)</a>
131-	135	F5.3	<a href="#">mag</a>	e_VTmag	[0.009, 1.468]? <a href="#">s.e. of VT (7)</a>
137-	139	I3	<a href="#">0.1arcsec</a>	prox	[3, 999] <a href="#">proximity indicator (8)</a>
	141	A1	---	TYC	[T] <a href="#">Tycho-1 star (9)</a>
143-	148	I6	---	HIP	[1, 120404]? Hipparcos number
149-	151	A3	---	CCDM	<a href="#">CCDM component identifier for HIP stars(10)</a>
153-	164	F12.8	<a href="#">deg</a>	RAdeg	Observed Tycho-2 Right Ascension, ICRS
166-	177	F12.8	<a href="#">deg</a>	DEdeg	Observed Tycho-2 Declination, ICRS
179-	182	F4.2	<a href="#">yr</a>	EpRA-1990	[0.81, 2.13] epoch-1990 of RAdeg

184-187	F4. 2	<a href="#">yr</a>	EpDE-1990	[0. 72, 2. 36]	epoch-1990 of DEdeg
189-193	F5. 1	<a href="#">mas</a>	e_RAdeg	<a href="#">s. e. RA*cos(dec), of observed Tycho-2 RA (5)</a>	
195-199	F5. 1	<a href="#">mas</a>	e_DEdeg	<a href="#">s. e. of observed Tycho-2 Dec (5)</a>	
201	A1	---	posflg	[ DP]	<a href="#">type of Tycho-2 solution (11)</a>
203-206	F4. 1	---	corr	[-1, 1]	correlation (RAdeg, DEdeg)

表 3.5: Tycho-2 星表(I/259)的数据说明<sup>[11]</sup>

### 3.2.4 星表数据库的入库规则:

由 CDS 星表文件导入到数据库中需要制定统一的, 正确的入库规则, 才能不丢失和更改星表数据的信息, 提高星表的查询效率。

#### 1. 数据库名称的创建规则:

星表所在的数据库名称就是和 CDS 星表的最上层目录一样, 比如 Tycho-2 星表(I/259)在数据库 I 下。对于某些大型星表要单独创建数据库, 数据库名为星表的常用缩写名称。

#### 2. 星表数据文件到表的映射规则和表名设置规则:

需要按照星表数据的总大小来分别对待, 绝大部分星表数据文件总量较小, 一般将这些星表数据文件导入到一个表中不会超过 2G, 不会出现问题; 但是少数星表很大, 一个表不能容下, 那就按照赤纬将数据按天区分布分成若干个“条”, 每一条包含赤纬在某个值范围内的所有天体数据, 这个分割要保证入库后表小于 2G。比如将 2MASS 分割成 92 个不等宽的“条”, 92 个表, 每个表大小一般为 1.27GB, 而将 USNOB 1.0 星表按天区分割成 180 个等宽“条”, 180 个表, 每个表大小不一致, 最大为 1.8GB。

小星表导入 MySQL 表名就是将这个星表的 CDS 路径名称去掉“/”符号, 如 Tycho-2 (I/259)星表的 MySQL 中表的名称为 I259;

大星表一般存储在单独的数据库中, 一般表名与数据文件的名称相同。

#### 3. MySQL 表中数据字段名的设置规则:

数据字段(属性列)的名称设置规则: MySQL表中数据字段名一般按照星表说明给出的列名, 允许出现大小写字母, 但是要对下列字符做一些转换, 包括

要将“/”、“.”、“-”、“'”“””转换成下划线“\_”。不允许字段名与MySQL数据库的保留字一样，若星表说明给出的列名与MySQL数据库的保留字一样，则在后面补一个下划线“\_”。MySQL保留字参见[http://www.phpe.net/mysql\\_manual/06-1.html#Legal\\_names](http://www.phpe.net/mysql_manual/06-1.html#Legal_names)，例如，一个最经常在说明中给出的列名dec或DEC(一般用来表示赤经)，这是在MySQL字段名中不允许出现的，要用dec\_或DEC\_。

另，没有具体意义的字段最好不导入MySQL表中。

#### 4. 数据字段（属性列）的数据类型的映射规则：

数据表说明中给出了第二列就是数据类型说明，它与MySQL定义的列的数据类型不同，我们制定了一个简单的易于操作的类型映射关系：

CDS 给出的数据类型	映射条件	MYSQL 数据类型
I[M]	$M \leq 2$	TINYINT(M)
	$2 < M \leq 4$	SMALLINT(M)
	$4 < M \leq 6$	MEDIUMINT(M)
	$6 < M \leq 9$	INT(M)
	$9 < M$	BIGINT(M)
F[M,N]	$M > 8$	DOUBLE(M,N)
E[M,N]	$M \leq 8$	FLOAT(M,N)
D[M,N]		
A[M]	$M = 1$	CHAR(1)
	$1 < M \leq 255$ 数据不定长	VARCHAR(M)
	$1 < M \leq 255$ 数据定长或 空格有意义	CHAR(M)
	$M > 255$	TEXT

表 3.6: CDS 给出的数据类型说明到 MYSQL 字段数据类型的映射

#### 5. 数据字段是否为空的设置规则：

当数据文件的某数据字段出现缺值时，将本字段设置允许为空。如没有任何

缺值，则置为非空。

#### 6. 赤经和赤纬列的添加规则：

在 CDS 星表中，大部分星表都用 7 个字段表示赤经和赤纬信息，即采用 60 进制方式的时、角、分、正负号、度、分、秒（一般说明中用 RAh, RAm, RAs, DE-, DEd, DEm, DEs 作为列名）7 个列，这不利于星表的位置查询，所以需要在表的末尾增加两列，名称不能和表的其他字段相同，可以设为 Radeg 和 Dedeg，或者按照其历元设为 RAJ2000, DEJ2000（当历元为 J2000 时），分别表示以十进制度为单位的赤经赤纬。

类型设置规则为：

赤经的小数点后的位数=RAs 的小数点后的位数+4；

赤纬的小数点后的位数=DEs 的小数点后的位数+5；

赤经的宽度=赤经的小数点后的位数+4；

赤纬的宽度=赤纬的小数点后的位数+4；

赤经，赤纬的数据类型：宽度>7 时，为 Double，否则为 Float（一般不会出现）；

如：RAs 类型为 E4,2 时，赤经的类型为 Double(10,6)，

DEs 类型为 E3,1 时，赤纬的类型为 Double(10,6)，

赋值规则为：

赤经 = (RAh+RAm/60.0+ RAs/3600.0)×15.0

$$\text{赤纬} = \begin{cases} (\text{DEd} + \text{DEm}/60.0 + \text{Des}/3600.0) \times 1.0 & (\text{当 DE-} = "+" \text{时}) \\ (\text{DEd} + \text{DEm}/60.0 + \text{Des}/3600.0) \times -1.0 & (\text{当 DE-} = "-" \text{时}) \end{cases}$$

#### 7. 索引的添加规则：

要为星表内表示天体名称和 ID 号的列添加单列索引。也必须要为星表中表示位置信息的列添加单列索引，比如表示赤道天体坐标位置的赤经、赤纬列，以及银经银纬等，如果表内有多个历元的赤经、赤纬，也必须为他们都添加单列索引。

由于 MySQL 只支持 B-Tree，不支持 R-Tree，所以不添加基于两列（比如赤经、赤纬列）的索引，这样会影响到查询性能。

#### 8. 表的权限的设定规则：



在数据导入 MySQL 表，并设定完毕字段是否为空，和添加完赤经和赤纬列和必要的索引后，将普通用户对表的权限设定为 Select，不允许对表进行 delete，insert 和 update 操作。

将 Tycho-2 星表(I/259)按照入库规则导入到星表数据库中后的结构：

字段名称	类型	索引	主键	是否为空	说明
TYC1	smallint(4)			N	
TYC2	mediumint(5)			N	
TYC3	tinyint(1)			N	
pflag	char(1)			N	
RAmdeg	Double(12,8)			Y	
DEmdeg	Double(12,8)			Y	
pmRA	float(7,1)			Y	
pmDE	float(7,1)			Y	
e_RAmdeg	smallint(3)			Y	
e_DEmdeg	smallint(3)			Y	
e_pmRA	float(4,1)			Y	
e_pmDE	float(4,1)			Y	
EpRAm	float(7,2)			Y	
EpDEm	float(7,2)			Y	
Num	tinyint(2)			Y	
q_RAmdeg	float(3,1)			Y	
q_DEmdeg	float(3,1)			Y	
q_pmRA	float(3,1)			Y	
q_pmDE	float(3,1)			Y	
BTmag	float(6,3)			Y	
e_BTmag	float(5,3)			Y	
VTmag	float(6,3)			Y	
e_VTmag	float(5,3)			Y	
prox	smallint(3)			N	
TYC	char(1)			N	
HIP	mediumint(6)			Y	
CCDM	char(3)			N	
RAdeg	Double(12,8)		Y	N	
DEdeg	Double(12,8)		Y	N	
EpRA_1990	float(4,2)			N	
EpDE_1990	float(4,2)			N	
e_RAdeg	float(5,1)			N	
e_DEdeg	float(5,1)			N	
posflg	char(1)			N	
corr	float(4,1)			N	

表 3.7: Tycho-2 星表入库的表结构

### 3.2.5 星表元数据库的设计与实现:

#### 3.2.5.1 星表元数据库概述:

传统的天文数据库缺乏元数据信息的管理,我们基于目前的资源元数据标准,以及 VOTable 数据标准对列(FIELD)的描述格式,并结合 CDS 星表的星表说明标准和星表的一般属性,我们设计了一个能对以下信息进行良好管理的星表元数据库:

- (1) 星表资源的元数据信息,描述结构基本符合 IVOA 的资源元数据标准 1.0 版(Resource Metadata 1.0);
- (2) 星表字段(列)的元数据信息,比如 UCD,物理单位,详细说明等等,描述结构内容与 VOTable 标准中 FIELD 元素结构基本一致;对 UCD 信息的描述部分,包含 UCD1 和 UCD1+词汇表,以及两者之间的对应转换关系。
- (3) 星表与星表表数据库中的表之间的对应关系。

#### 3.2.5.2 星表元数据库的命名规则:

- (1) 数据库名为 VO,表名和字段名均由大写字母或下划线组成,不允许出现小写字母或其他字符;
- (2) 每一个名字的第一个元素一定要是大写字母,对象名最长限度为 31 个字符;
- (3) 保证表名的唯一性。在星表元数据库 VO 范围内的表名称必须是唯一的。

#### 3.2.5.3 星表元数据库的逻辑设计:

数据实体—关系图:如图 3.3。

数据实体描述:

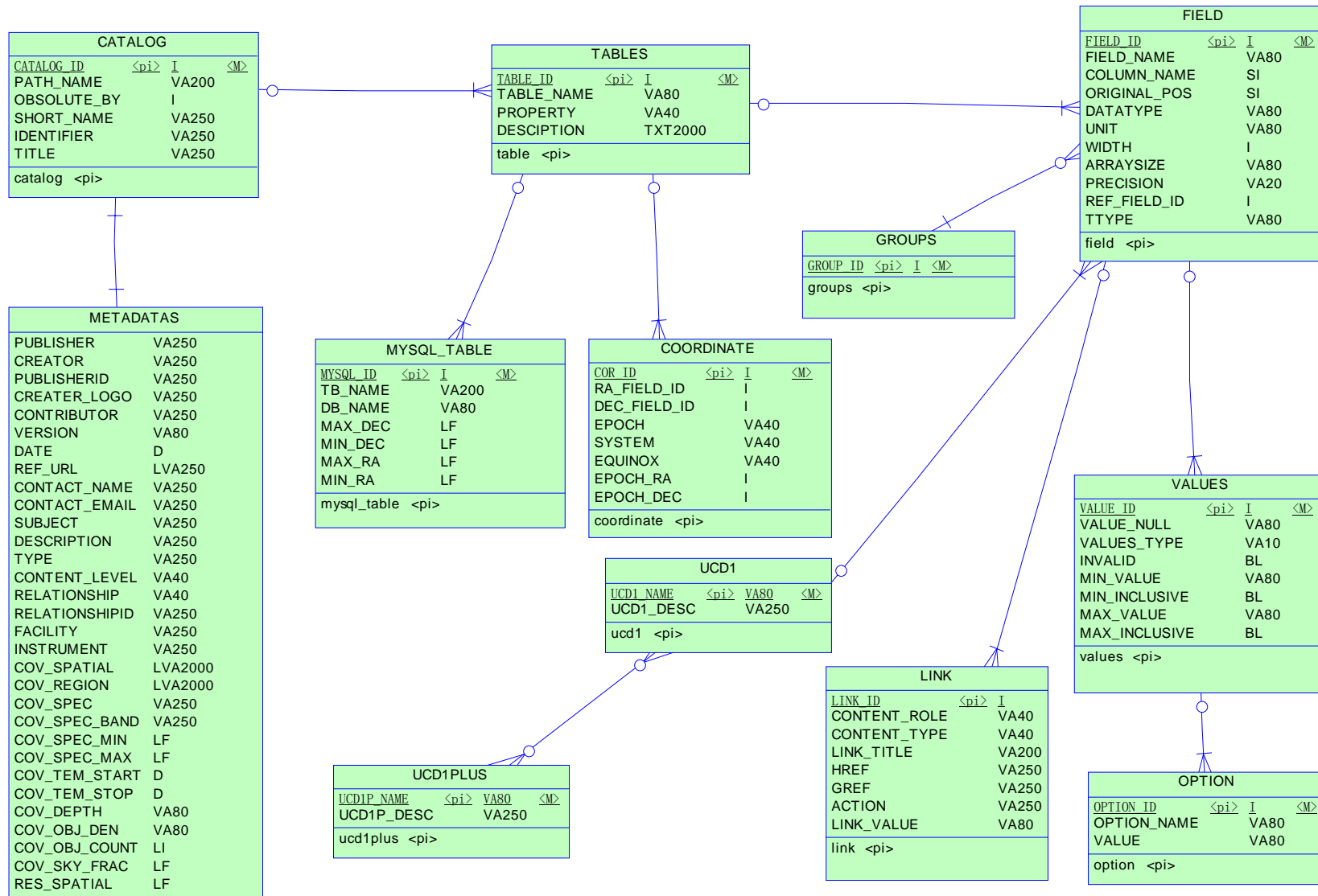


图 3.3: 星表元数据库逻辑设计实体-关系图

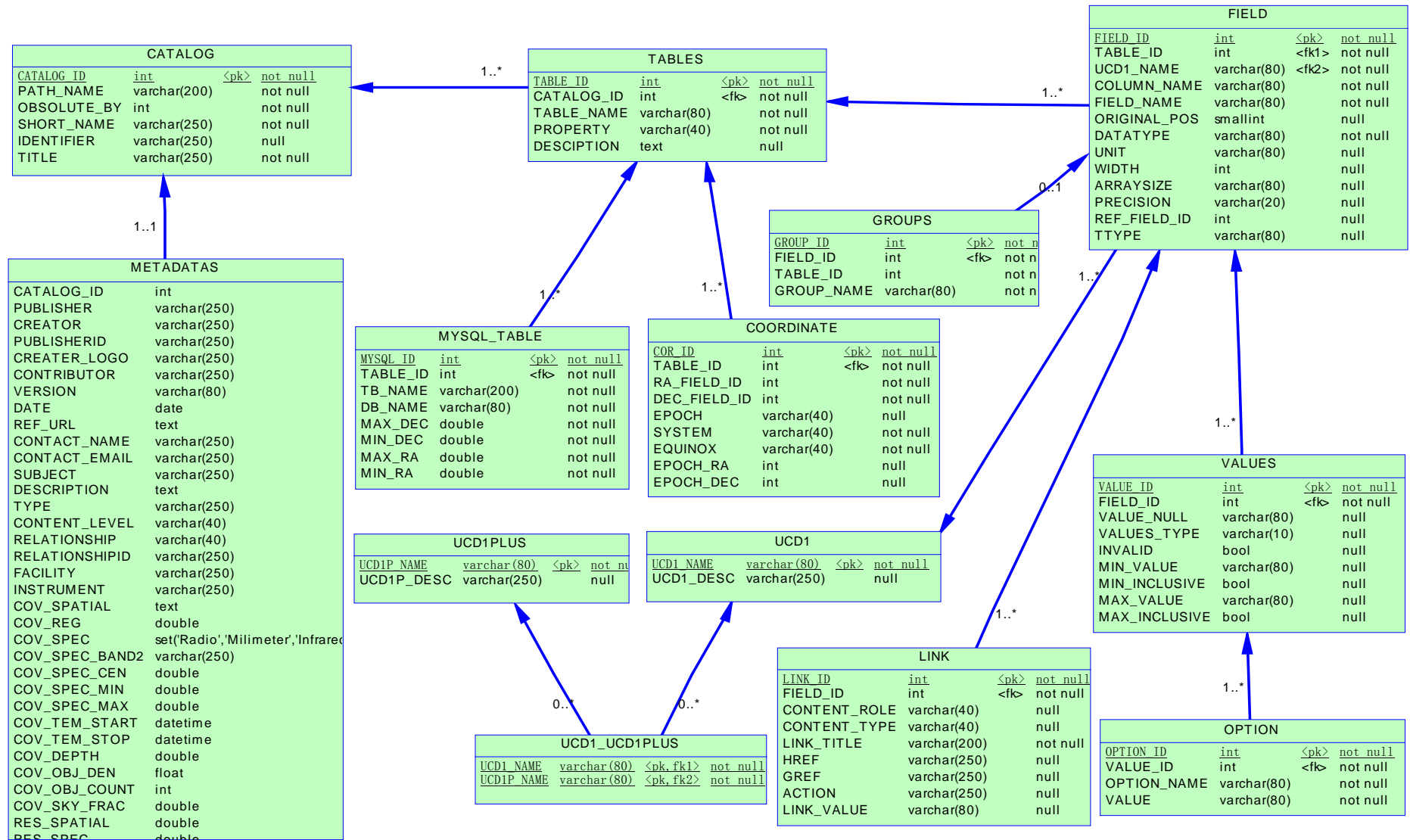


图 3.4: 星表元数据库的物理结构图

功能	实体名	数据库表名	实体描述
	CATALOG	CATALOG	星表的基本信息
星表资源 元数据	METADATA	METADATA	星表资源元数据信息，内容按照 IVOA 的资源元数据标准 1.0
	TABLES	TABLES	星表包含的数据表的基本信息，一个星表可能包含多个数据表
	MYSQL_TABLE	MYSQL_TABLE	数据表对应的 MySQL 数据库中的表的基本信息，一个数据表可能存储在 MySQL 多个表中
列元数据	COORDINATE	COORDINATE	星表坐标系属性和在数据表中列对应的列
	GROUPS	GROUPS	
	FIELD	FIELD	数据表中列的信息汇总，内容遵照 VOTable 1.0 格式标准 FIELD 元素
	LINK	LINK	数据表中列的信息链结，内容遵照 VOTable 1.0 格式标准 FIELD 元素的子元素 LINK
	VALUES	VALUES	数据表中列的取值范围的信息，内容遵照 VOTable 1.0 格式标准 VALUE 元素
	OPTION	OPTION	数据表中列的允许取的值，内容遵照 VOTable 1.0 格式标准 OPTION 元素
UCD 词 汇	UCD1	UCD1	UCD1 词汇表
	UCD1PLUS	UCD1PLUS	UCD1+词汇表
		UCD1_UCD1PLUS	UCD1 词汇表 UCD1+词汇之间的映射关系

表 3.8: 数据实体描述

实体关系描述:

实体关系		关系描述
CATALOG: METADATA	1: 1	星表必须而且只有一个资源元数据信息, 某个资源元数据信息必须属于某一个星表
CATALOG: TABLES	1: n	一个星表可以包含零到多个数据表, 一个数据表必须属于某一个星表
TABLES: MYSQL_TABLE	1: n	一个数据表可能存储在多个 MYSQL 表中, 一个 MYSQL 表属于某个数据表
TABLES: COORDINATE	1: n	一个数据表可以有多个坐标信息, 坐标信息必须属于一个数据表
TABLES: FIELD	1: n	一个数据表有多个列, 一个列必须属于一个表
FIELD: LINK	1: n	一个列的基本信息可以包括多个链接, 一个链接必须属于一个列的信息
FIELD: VALUES	1: n	一个列的基本信息可以包括多个取值的描述
VALUES: OPTION	1: n	一个取值描述可以包括多个允许取的值, 一个允许取的值必须属于一个取值描述
FIELD: UCD1	n: 1	一个 UCD1 可以用于多个列的描述, 而一个列的 UCD 描述必须来自 UCD1
FIELD: GROUPS	n: 0	一个组由一到多个列组成, 一个列可以不属于一个组
UCD1: UCD1PLUS	m: n	UCD1 和 UCD1+之间的转换对应关系

表 3.9 实体关系描述

### 3.2.5.4 星表元数据库的物理设计:

物理设计: 采用 MySQL 数据库系统的星表元数据库物理结构如图 3.4。

数据库表汇总介绍:

#### 1. 表: CATALOG

字段名称	类型	索引	主键	是否为空	说明
CATALOG_ID	int		Y	N	星表 ID
PATH_NAME	varchar(200)			N	CDS 星表路径
OBSOLUTE_BY	int			N	取代本星表的其他星表的 ID
SHORT_NAME	varchar(250)			N	星表缩写名称
IDENTIFIER	varchar(250)			Y	IVOA ID
TITLE	varchar(250)			N	星表正式名称(全称)

## 2. 表: METADATA

字段名称	类型	索引	主键	是否为空	说明
CATALOG_ID	int	Y		N	星表 ID
PUBLISHER	varchar(250)			Y	星表发布者
CREATOR	varchar(250)			Y	星表创建者
PUBLISHERID	varchar(250)			Y	星表发布者 ID
CREATER_LOGO	varchar(250)			Y	星表创建者
CONTRIBUTOR	varchar(250)			Y	星表贡献者
VERSION	varchar(80)			Y	星表版本
DATE	data			Y	发布日期
REF_URL	text			Y	参考链接
CONTACT_NAME	varchar(250)			Y	联系人名称 (Sang Jian)
CONTACT_EMAIL	varchar(250)			Y	联系人 Email(sang@lamoost.org)
SUBJECT	varchar(250)			N	关键字列表
DESCRIPTION	text			Y	星表详细说明
TYPE	varchar(250)			N	星表类型, 一般为'Catalog'
CONTENT_LEVEL	varchar(40)			N	内容层次描述, 面向的用户。
RELATIONSHIP	varchar(40)			Y	与其他资源的关系
RELATIONSHIPID	varchar(250)			Y	有关系的资源的 IVOA ID

FACILITY	varchar(250)	N	观测单位机构, 列表
INSTRUMENT	varchar(250)	N	专门的观测设备的名称, 列表
COV_SPATIAL	text	N	观测覆盖的天区
COV_REG	double	Y	
COV_SPEC	set('Radio','Millimeter', , 'Infrared','Optical','Ultraviolet','X-ray','Gamma-ray')	Y	观测波段, 可取值依次为射电、微波、红外、光学、紫外、X射线和 $\gamma$ 射线。
COV_SPEC_BAND	varchar(250)	Y	覆盖波段, 列表
COV_SPEC_CEN	double	Y	观测波段中心波长
COV_SPEC_MIN	double	Y	观测波段最小波长, 单位: 米
COV_SPEC_MAX	double	Y	观测波段最大波长, 单位: 米
COV_TEM_STRAT	datetime	Y	观测开始时间, 一般精确到日期
COV_TEM_STOP	datetime	Y	观测结束时间, 一般精确到日期
COV_DEPTH	double	Y	深度覆盖(灵敏度), 单位: J <sub>r</sub>
COV_OBJ_DEN	float	Y	观测目标密度, 以每平方度计数
COV_OBJ_COUNT	int	N	总观测目标数目, 即星表行数
COV_SKY_FRAC	double	Y	覆盖天区的分数, (0, 1]
RES_SPATIAL	double	Y	空间(角)分辨率, 单位: 十进制度
RES_SPEC	double	Y	光谱分辨率, $\Delta\lambda/\lambda$ 比率值
RES_TEMP	double	Y	时间分辨率, 单位: 秒
RIGHTS	varchar(40)	Y	权利
DATA_QUALITY	enum('A','B','C','U')	Y	数据质量说明
UNCER_PHOTO	double	Y	光度测量不确定性, 单位: J <sub>r</sub>
UNCER_SPATIAL	double	Y	位置测量不确定性, 单位: 度
UNCER_SPEC	double	Y	波长不确定性, 单位: 米
UNCER_TEMP	double	Y	时间不确定性, 单位: 秒



## 3. 表: TABLES

字段名称	类型	索引	主键	是否为空	说明
TABLE_ID	int		Y	N	数据表 ID
CATALOG_ID	int			N	星表 ID
TABLE_NAME	varchar(80)			N	数据表名称
PROPERTY	varchar(40)			N	数据表属性
DESCRIPTION	text			Y	数据表说明

## 4. 表: MYSQL\_TABLE

字段名称	类型	索引	主键	是否为空	说明
MYSQL_ID	int		Y	N	
TABLE_ID	int			N	对应的数据表的 ID
TB_NAME	varchar(200)			N	MYSQL 数据库中表的名称
DB_NAME	varchar(80)			N	数据库表所在数据库名称
MAX_DEC	double			N	表中赤纬最大取值
MIN_DEC	double			N	表中赤纬最小取值
MAX_RA	double			N	表中赤经最大取值
MIN_RA	double			N	表中赤经最小取值

## 5. 表: COORDINATE

字段名称	类型	索引	主键	是否为空	说明
COR_ID	int		Y	N	ID
TABLE_ID	int			N	对应数据表的 ID

RA_FIELD_ID	int	N	赤经所在数据列 ID
DEC_FIELD_ID	int	N	赤纬所在数据列 ID
EPOCH	varchar(40)	Y	观测历元
SYSTEM	varchar(40)	N	坐标系统
EQUINOX	varchar(40)	N	坐标系平春分点历元
EPOCH_RA	int	Y	赤经观测历元
EPOCH_DEC	int	Y	赤纬观测历元

6. 表: GROUPS

字段名称	类型	索引	主键	是否为空	说明
GROUP_ID	int		Y	N	组 ID
FIELD_ID	int			N	数据列 ID
TABLE_ID	int			N	表 ID

7. 表: FIELD

字段名称	类型	索引	主键	是否为空	说明
FIELD_ID	int		Y	N	数据列 ID
TABLE_ID	int			N	所在的表的 ID
UCD1_NAME	varchar(80)			N	数据列的 UCD
FIELD_NAME	varchar(80)			N	数据列的名称
COLUMN_NAME	varchar(80)			N	数据列对应的 MySQL 表中的列名
ORIGINAL_POS	smallint			Y	数据列在表中的位置
DATATYPE	varchar(80)			N	数据列的数据类型
UNIT	varchar(80)			Y	数据列单位

WIDTH	int	Y	数据列宽度
ARRAYSIZE	varchar(80)	Y	数据列为数组时，数组各维的大小
PRECISION	varchar(20)	Y	数据列为浮点型时，其精度
REF_FIELD_ID	int	Y	参考其他数据列的 ID
TTYPER	varchar(80)	Y	

8. 表: LINK

字段名称	类型	索引	主键	是否为空	说明
LINK_ID	int		Y	N	ID
FIELD_ID	int			N	对应列的 ID
CONTENT_ROLE	varchar(40)			Y	内容角色
CONTENT_TYPE	varchar(40)			Y	内容类型
LINK_TITLE	varchar(200)			N	链结主题名称
HREF	varchar(250)			Y	href
GREF	varchar(250)			Y	gref
ACTION	varchar(250)			Y	
LINK_VALUE	varchar(80)			Y	链结值

9. 表: VALUES

字段名称	类型	索引	主键	是否为空	说明
VALUE_ID	int		Y	N	ID
FIELD_ID	int			N	数据列 ID
VALUE_NULL	varchar(80)			Y	用何表示值为空
VALUES_TYPE	varchar(10)			Y	值的类型

INVALID	bool	Y	
MIN_VALUE	vchar(80)	Y	数据列最小允许取值
MIN_INCLUSIVE	bool	Y	有效值是否包括上述最小值
MAX_VALUE	vchar(80)	Y	数据列最大允许取值
MAX_INCLUSIVE	bool	Y	有效值是否包括上述最大值

10. 表: OPTION

字段名称	类型	索引	主键	是否 为空	说明
OPTION_ID	int		Y	N	ID
VALUE_ID	int			N	值 ID, VALUE 表
OPTION_NAME	vchar(80)			N	可选取值的名称意义
VALUE	vchar(80)			N	可选取值

11. 表: UCD1

字段名称	类型	索引	主键	是否 为空	说明
UCD1_NAME_	vchar(80)		Y	N	UCD1 名称
UCD1_DESC	vchar(250)			Y	UCD1 的意义说明

12. 表: UCD1PLUS

字段名称	类型	索引	主键	是否 为空	说明
UCD1P_NAME	vchar(80)		Y	N	UCD1+名称
UCD1P_DESC	vchar(250)			Y	UCD1+意义说明

## 13. 表: UCD1\_UCD1PLUS

字段名称	类型	索引	主键	是否 为空	说明
UCD1_NAME	varchar(80)		Y	N	UCD1 名称
UCD1P_NAME	varchar(80)		Y	N	UCD1+名称

## 3.2.5.5 数据库安全性设计

用户角色的权限如下:

角色	可以访问的表	操作权限
数据库管理员(root)	ALL	ALL
数据信息录入员(oper)	ALL	INSERT, DELETE, UPDATE, SELECT
普通用户(user)	ALL	SELECT

表 3.10: 用户角色及权限

## 3.2.6 星表数据入库程序的实现

目前为了减少数据库入库的工作量,我们利用 Java 语言开发了一个将星表自动导入星表数据库的程序,完全实现了星表数据自动导入星表数据库,并且也将有关的列元数据导入星表元数据库。如图 3.5:

最复杂的工作,也是最繁重的工作是元数据入库工作,由于原始的 CDS 星表 ReadMe 的格式为文本格式,结构复杂且不完全统一,从说明中自动提取有关星表的元数据信息是根本不可能实现的。所以元数据的收集整理工作主要由数据库录入员手工进行。

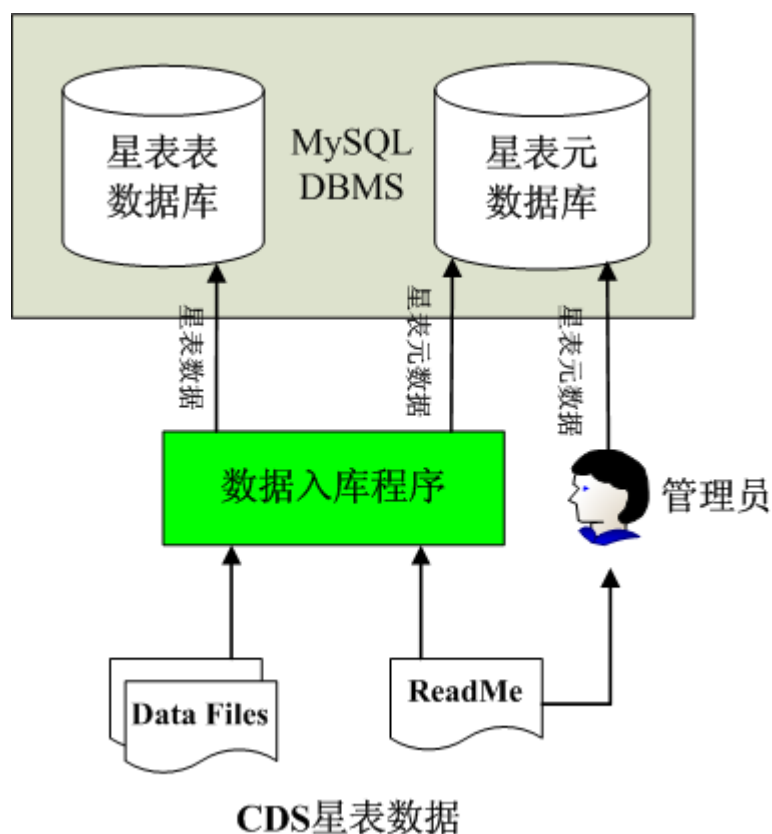


图 3.5：星表数据库数据入库机制

### 星表数据入库的流程：

数据入库程序的流程如图所示：左边为进行的工作，右边为实现其功能的自动入库程序中的 Java 类。

1. 解析 ReadMe 文件，得到每一列的起止字节，类型，单位，名称和注释，起止字节和类型用来读取数据文件，生成插入语句；类型和名称用来生成创建表的语句；而单位，名称和注释作为列的元数据信息的一部分被导入到星表元数据库中；
2. 根据解析生成的类型和名称映射成 MySQL 表中的列的数据类型和列名。生成创建表的语句，执行这个语句生成表；
3. 根据解析提供的每一列的起止字节和类型，读取解析数据文件，将每一行数据构建成插入语句，并向 2 生成的表插入数据，直至所有数据文件的行被导入数据库表中；

4. 将数据库表中没有意义的列删除，多数星表不需要这一步；
5. 将表中没有出现缺值的列改为非空。在开始创建表时默认每一列都可以为空，在执行插入操作时将缺值项设为空。之所以采用入库后再将列修改为非空，而不是在开始创建表时就设定好，是因为 ReadMe 没有全部提供这一列是否为空的信息。如果按照非空创建表，在步骤 3 插入操作中遇到空值就会抛出错误。如果采用插入前先探测一遍数据的方法会花费更多的时间。所以只能采用入库后再修改的方式；
6. 大部分给出的坐标值（赤经，赤纬）是用 7 个参数来表示，即采用 60 进制的时、角、分、正负号、度、角分、角秒（RAh, RAm, RAs, DE-, DEd, DEm, DEs）的描述格式，这样不利于进行查询，我们需要在表的尾部增加两列，来表示 10 进制单位为度的赤经，赤纬，它们的值由那 7 列换算出；
7. 要给星表中可能常用做查询条件的列添加索引，以提高检索数据效率。DecorateMySQLTable.class 中的函数实现了上述 4,5,6,7 步；
8. 将解析 ReadMe 文件得到的列的单位，名称，注释信息和数据库表的属性信息作为元数据插入到列元数据表中。

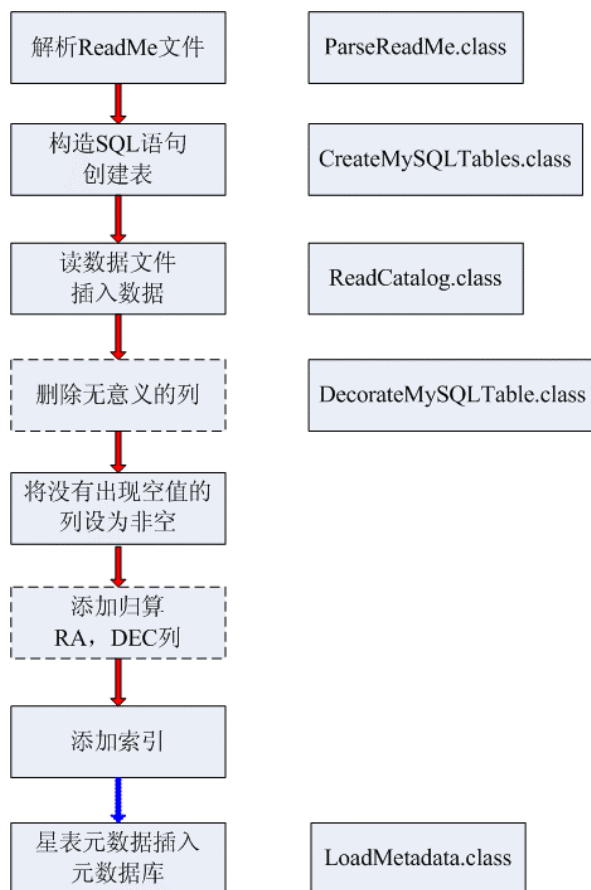


图 3.6: 星表入库流程图

### 3.2.7 目前已经入库的星表列表:

波段	星表名称	CDS	目标数目	数据量
可见光	UCAC2-BSS	I/294	430000	65MB
	UCAC2	I/289	48330571	4.9GB
	SPM3.2		10764905	998MB
	USNO-B1.0	I/284	1045913669	169GB
	ASCC-2.5	I/280	2501313	309MB
	AC2000.2	I/275	4621751	414MB
	M2000	I/272	2286322	87MB
	Tycho-2.	I/259	2539913	537MB



	GSC1.2	I/254	25241730	1.3GB
	Hipparcos	I/239	118218	31MB
	SAO	I/131A	258997	52MB
	RC3	VII/155	23011	7MB
红外	2MASS-PSC	II/246	470992970	134GB
	2MASS-ESC	VII/233	1647599	2.3GB
	IRAS-F	II/156A	173044	33MB
射电	FIRST	VIII/71	811117	111MB
	NVSS	VIII/65	1773484	266MB
X-射线	RASS-FSC	IX/29	105924	13MB
	ROSHRI	IX/28A	108514	16MB
	RASS-BSC	IX/10A	18806	3MB
统计	112 个表			311GB

表 3.11: 目前已经导入到天文星表数据库的星表

## 参考文献

- [1] [BADC] Beijing Astronomical Data Center. <http://badc.lamost.org>
- [2] [WDC] World Data Center System.  
<http://www.ngdc.noaa.gov/wdc/wdcmain.html>
- [3] [DR1] The SDSS Data Release 1. <http://www.sdss.org/dr1/start/aboutdr1.html>
- [4] [MySQL] <http://www.mysql.com>
- [5] MySQL 管理员指南. <http://art.055.cn/web/358.htm>
- [6] Jay Greemspan,Brad Bulger. MySQL/PHP Database Applications.电子工业出版社,2001 第一版
- [7] Paul DuBois. MySQL. 机械工业出版社,2000 年第一版
- [8] [PowerDesigner] [www.sybase.com/products/enterprisemodeling/powerdesigner](http://www.sybase.com/products/enterprisemodeling/powerdesigner)
- [9] [CDS] Centre de Données astronomiques de Strasbourg.  
<http://cdsweb.u-strasbg.fr/>
- [10] Astronomical catalogues and tables adopted standards .  
<http://vizier.u-strasbg.fr/doc/catsd.htm>
- [11] [Tycho-2] The Tycho-2 Catalogue. <http://cdsweb.u-strasbg.fr/viz-bin/Cat?I/259>



## 第四章 虚拟天文台的数据服务

### § 4.1 网格及相关技术的定义

#### 4.1.1 网格的定义

全球网格研究的领军人物、美国 Argonne 国家实验室的资深科学家、美国 Globus 项目的领导人 Ian Foster 曾在 1998 年出版的《The Grid: Blueprint for a New Computing Infrastructure》一书中这样描述网格：“网格是构筑在互联网上的一组新兴技术，它将高速互联网、高性能计算机、大型数据库、传感器、远程设备等融为一体，为科技人员和普通老百姓提供更多的资源、功能和交互性。互联网主要为人们提供电子邮件、网页浏览等通信功能，而网格功能则更多更强，让人们透明地使用计算、存储等其他资源。”

2000 年，Ian Foster 在《The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration》这篇论文中把网格进一步描述为“在动态变化的多个虚拟机构间共享资源和协同解决问题。”但是，人们仍然就什么是网格而争论不休。2002 年 7 月，Ian Foster 在《What is the Grid? A Three Point Checklist》<sup>[18]</sup>一文中，限定网格必须同时满足三个条件：

1. 在非集中控制的环境中协同使用资源；
2. 使用标准的、开放的和通用的协议和接口；
3. 提供非平凡的服务。

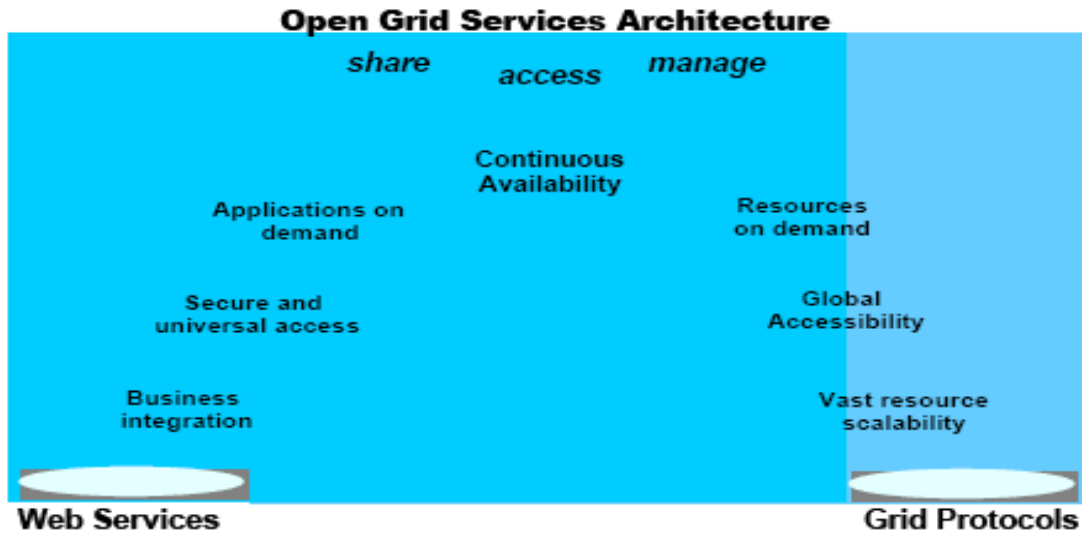
这三个条件非常严格，将 P2P<sup>[19]</sup>、SUN ONE Grid Engine<sup>[20]</sup>、Condor<sup>[21]</sup>、Entropia<sup>[22]</sup>、MultiCluster<sup>[23]</sup> 等都被排除在网格之外。

Ian Foster 把他头脑中的网格概念描绘清楚了，但并不是所有人都同意他的观点。例如，有许多人赞同广义的网格概念，把它称作超级全球网格 GGG (Great Global Grid)。GGG 不仅包括计算网格、数据网格、信息网格、知识网格、商业网格，还

包括一些已有的网络计算模式，例如对等计算P2P(Peer to Peer)、寄生计算等。可以这样认为，Ian Foster 赞成狭义的“网格观”，而GGG 是一种广义的“网格观”。

#### 4.1.2 OGSA (Open Grid Services Architecture)

是一个面向服务的体系结构。在 OGSA 中，一切都以 Grid 服务的形式体现。基于 Grid 和 WEB 服务的思想和技术,OGSA 体系定义了一个统一的对外服务语义,即定义了标准的瞬时 Grid 服务实例的创建、命名和发现机制;为服务实例提供了地域透明性和多协议绑定;并提供了与本地平台系统的集成机制;以 WSDL 接口和相关约束的格式定义了创建和组织高级分布式系统所需的机制,其中包括生命期管理、变动管理、通告等。



#### 4.1.3 GT3 (Globus Toolkits 3.0)

作为OGSA的一种实现，GT3的服务模块包括：核心服务、安全服务、基本服务、数据服务和其他服务。这几种服务之间的关系如图所示：



GT3核心服务包括:

- **Factory:** 工厂服务。通过它的CreateService操作可以创建新的网格服务，返回网格服务句柄（Grid Service Handle, GSH），维护所需的服务数据元素。一个GSH是一个URL，用来为一个网格服务实例命名。为了使用这个服务，一个GSH必须转换为一个网格服务参考（Grid Service Reference, GSR）。GSR描述一个客户如何能与一个网格服务实例通信。HandleMap接口允许一个客户把一个GSH映射到GSR。GSH仅代表名字，GSR包括了传输协议的绑定信息和数据编码格式。
- **NotificationSource:** 用来接受用户通告订阅
- **HandleResolver:** 句柄解析器
- **Registration:** 通过返回一组网格服务的GSH来支持服务发现
- **NotificationSubscription:** 用户通告订阅
- **NotificationSink:** 用来异步发送通知消息
- **GridService:** 所有网格服务都必须支持的接口

GT3的核心服务要实现:

- OGSI规范

- 公用接口和API（包括服务数据、通告、查询、状态管理等）
- 网格服务容器基础设施
- 开发与运行环境

GT3安全服务主要包括：

- 传输层：SSL
- 消息层：WS-Security, SOAP
- 安全会话服务
- 消息签名与加密
- 信任授权、访问控制
- 基于虚拟主机环境的增强型资源安全模型

GT3基本服务包括：

- 索引服务
- 元数据目录索引服务
- 服务数据提供者基础设施
- 可管理作业服务
- 资源管理
  - GRAM的实现（GRAM是Globus资源管理体系结构中最底层的部分，通过使用一组WSDL/OGSI客户端接口进行作业的提交、监视和终止操作，实现作业的远程运行等活动。）
- 虚拟主机环境和路由基础设施
- 文件流式化服务
- 可靠文件传输服务（GridFTP、可重获的网格服务、进展与重新启动监视）

GT3数据服务包括：

- 数据访问
- 数据备份
- 数据缓存
- 元数据目录和服务集的服务
- Schema变换
- 存储
- 数据库

GT3其他服务包括:

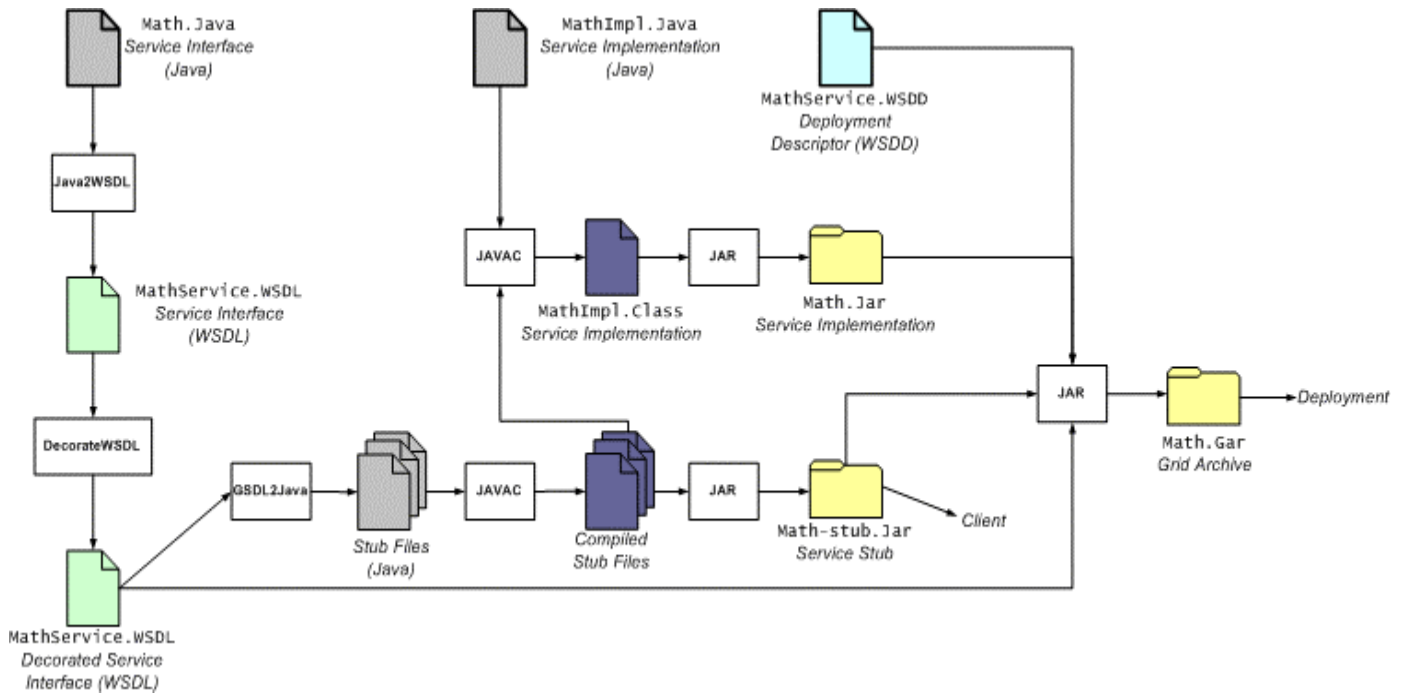
- 错误诊断
- 负载管理

#### 4.1.4 GT3 服务部署

中国虚拟天文台的数据服务都是采用基于 OGSA 的 GT3 部署实现的, 具体使用 GT3 部署服务的步骤在下例中做具体说明, 服务部署实现分为四个步骤:

- 1 提供一个服务接口
- 2 生成网格服务支持代码
- 3 实现服务
- 4 部署服务





以上图为例说明：

### 1 提供服务接口：

编写服务接口 `Math.java`。有两种方法可以用来提供网格的服务接口。一种是用 Java 语言编写一个接口，然后用工具生成 `WSDLPortType` 接口描述文件，以供用户远端的客户调用，这种方法可以容易、方便的生成 `WSDL` 文件，但是一些复杂的接口不是总能正确的转换为对应的 `WSDL` 文件；一种方法是自己编写 `WSDLPortType` 接口文件，可以完全控制服务接口的描述，适合于编写复杂的服务，但是直接书写 `WSDL` 对用户要求比较高。

### 2 生成网格服务支持代码：

这一步骤分为两步，生成 `WSDL` 文件和生成 `Stub`。编辑 `java` 接口文件，使用 Axis 工具 `Java2WSDL` 从 `java` 接口中产生 `WSDL` 文件 `MathService.wsdl`，再使用 GT3 的 `DecorateWSDL` 工具补充修饰，这样就得到了完整的 `WSDL` 文件；使用 GT3 工具 `GSDL2Java` 产生服务端和客户端所需要的所有 `Stub` 文件，负责 SOAP 协议和网络通信工作。

### 3 实现服务：

网格服务的实现仅仅是一个 `java` 类，我们称之为 `MathImpl.java`，其中提供了对

方法的实现。MathImpl.java 是 GridServiceImpl 的子类。所有的网格服务都必须从基类 GridServiceImpl 中衍生出来。同时，网格服务还实现了 MathPortType。PortType 是服务接口的另一种称呼，我们告诉 java 网格服务对外提供了一种特殊的接口：MathPortType。

#### 4 部署服务：

此时已经生成了网格服务所需的所有的 Java 代码：1).服务接口 2). WSDL 服务描述文件 3). stub 文件 4). 服务的实现。接着需要把所有的这些文件编译、归档，并部署到 host 环境中。

这一步需要 3 个步骤：

##### A. 编写服务部署描述文件

部署文件告诉 web 服务器应该怎样来发布网格服务。部署文件以 WSDD 的形式存在。

##### B. 建立 Gar 文件

编译所有文件，并打包。

##### C. 部署到主机环境

## § 4.2 星表 Cone-Search 服务的设计和实现

天文星表是天文研究中最常用到的数据,比如象现在 CDS 星表服务收录的天文星表已经达到 5 千多,实际数目要更多,总数据量也很大,现在很多数据中心都有大量的天文星表数据,实现这些分布式星表的统一访问具有非常重要的意义。天文星表同时也是经过良好处理、结构形式比较简单(二维表)的数据,也容易实现统一的访问。各国虚拟天文台都不约而同的首先对星表的统一访问进行了探讨。

首先对用户的查询需求进行分析,发现用户大部分都是对星表的小块天区进行查询,即 Cone Search—天区查询,查询天区的形状可以有多种多样,最常用的就是圆形和矩形,圆形对应的称为点圈天区查询(Radius Search),即给出查询中心点的赤道坐标,查询出给定半径内所有目标的信息,点圈天区查询也是 Cone Search 中使用最多的;矩形对应的称矩形天区查询(Rectangle Search),即给出查询中心点赤道坐标,给定在赤经经线方向和赤纬纬线方向上的张角大小,查询出对应的矩形天区内所有目标的信息。实际上,绝大部分现有的数据中心提供的星表数据查询都是这两种简单的天区查询,包括最有名的 CDS Vizier 服务,象 NED 的位置查询服务提供的也是这种查询。

美国虚拟天文台首先提出 Cone Search 统一服务接口,实际上概念有些混淆,他们指的是点圈天区查询(Radius Search),接口定义是基于 HTTP 协议的。我们也采纳了这个接口的定义,并用于网格服务。

### 4.2.1 星表 Cone Search 服务的创建规则

将每一个数据资源例如星表按照定义的 Cone Search 服务接口封装成一个数据服务。不能将多个星表封装在一个服务里,即使他们存储在同一台主机的同一个数据库管理系统中,由于很多数据中心主机内存存储多个数据集,所以一台主机可以提供多个数据服务,由于这些服务的地址名称不同,所以同一台主机上的数据服务间也不会发生冲突。同样,如果一个数据集分布星表在多台主机内,那么也只能基于多台主机部署一个数据服务。如下图所示:

### 4.2.2 星表 Cone Search 服务接口

星表 Cone Search 服务由接口 ConeSearchPortType 定义，其特定的方法只有一个，具体描述如下：

- **getVOTable**

1. 方法原型：byte[] getVOTable (double ra, double dec, double radius)
2. 功能：即上面提到的点圈天区查询，查询的天区为圆形，给出天区中心点坐标和天区半径，返回查询天区内所有的天体的信息。
3. 参数：接口仅有三个输入值，具体意义描述如下表所示：

参数名称	数据类型	参数意义	单位
ra	double	天区中心点赤经值	度
dec	double	天区中心点赤纬值	度
radius	double	搜索天区半径	度

表：getVOTable 方法的输入参数

4. 返回值：若成功，返回一 VOTable 格式的查询结果文档，这一文档必须是格式有效的，其中包含了所有符合查询要求的数据和星表的元数据信息。

失败—抛出异常，若输入参数非法，抛出 `ErrorInputParametersException` 异常，出现其他错误抛出 `ServiceInternalErrorException`。

注：抛出 `ErrorInputParametersException` 条件：当出现输入参数不在以下合法范围内时

- A)  $0.0 \leq ra < 360.0$ ;
- B)  $-90.0 \leq dec \leq +90.0$ ;
- C)  $0.0 \leq radius \leq 5.0$ ;

天文数据库查询服务

服务封装的单位还是资源，外延要大，包括所有的星表数据资源，以及所有存储在

关系型数据库中的表列数据资源，比如 SkyServer DR1 等。

### 4.2.3 基于 MySQL 天文星表数据库的 Cone-Search 服务的实现

### 4.2.4 基于文件系统的星表 Cone-Search 服务的实现

管理象星表这样的表列数据，传统的关系数据库软件提供了强大功能，所以现在大部分数据中心将表列数据存储的关系数据库管理系统中。我们基于第三章提到的天文星表数据库，实现了多个星表的 Cone-Search 服务。

#### 4.2.4.1 以数据文件+查询小程序释放的星表介绍：

部分大天文星表在被释放时，为了方便天文学家的使用，采用专门的文件格式存储星表数据，并且提供了一个查询程序，这些查询程序用 C 语言编写，一般仅仅提供简单的点圈天区和矩形天区查询功能，天文学家得到拷贝后仅需要简单的设置就可以在本地机器上查询这些星表的数据。

目前我们收集到这样的星表一共有 11 个，它们中包括了目前目标数目最大的几个星表，如 USNO B 1.0, 2MASS PSC, GSC2.2 等，这些星表已经成为我们天文数据库的一部分，并且我们基于它们搭建了网格环境下的星表 Cone-Search 服务，采用这些去搭建网格星表数据服务，这在一个首创性的工作。

详细的列表如下，下表包括了这些星表名称、提供的查询功能和获取方式，如果用户有兴趣，可以按照表中的地址下载这些星表的数据和查询程序。在“提供的查询功能”一栏中，RAD 表示提供了点圈天区查询功能，RECT 表示提供了矩形天区查询功能。

星表名称	CDS	目标个数	提供的 查询功能	获取地址
2MASS	<a href="#">II/246</a>	470,992,970	RAD, RECT	<a href="ftp://data.bao.ac.cn/cats/bincats/2MASS/">ftp://data.bao.ac.cn/cats/bincats/2MASS/</a>

APM	<a href="#">I/267</a>	166,466,987	RAD	<a href="ftp://data.bao.ac.cn/cats/bincats/APM/">ftp://data.bao.ac.cn/cats/bincats/APM/</a>
DENIS	<a href="#">B/denis</a>	195,204,157	RAD	<a href="ftp://data.bao.ac.cn/cats/bincats/DENIS/">ftp://data.bao.ac.cn/cats/bincats/DENIS/</a>
GSC1.1	<a href="#">I/220</a>	25,258,765	RAD, RECT	<a href="ftp://data.bao.ac.cn/cats/I/220/GSC/">ftp://data.bao.ac.cn/cats/I/220/GSC/</a>
GSC1.2	<a href="#">I/254</a>	25,241,730	RAD, RECT	<a href="ftp://data.bao.ac.cn/cats/bincats/GSC_1.2">ftp://data.bao.ac.cn/cats/bincats/GSC_1.2</a>
GSC ACT	<a href="#">I/255</a>	25,241,730	RAD, RECT	<a href="ftp://data.bao.ac.cn/cats/bincats/GSC_ACT">ftp://data.bao.ac.cn/cats/bincats/GSC_ACT</a>
GSC2.2	<a href="#">I/271</a>	455,851,237	RAD, RECT	<a href="ftp://data.bao.ac.cn/cats/bincats/GSC2.2">ftp://data.bao.ac.cn/cats/bincats/GSC2.2</a>
UCAC1	<a href="#">I/268</a>	27,425,433	RAD, RECT	<a href="ftp://data.bao.ac.cn/cats/bincats/UCAC1">ftp://data.bao.ac.cn/cats/bincats/UCAC1</a>
UCAC2	<a href="#">I/289</a>	48,330,571	RAD, RECT	<a href="ftp://data.bao.ac.cn/cats/bincats/UCAC2">ftp://data.bao.ac.cn/cats/bincats/UCAC2</a>
USNO A2.0	<a href="#">I/252</a>	526,280,881	RAD, RECT	<a href="ftp://data.bao.ac.cn/cats/bincats/USNO_A2">ftp://data.bao.ac.cn/cats/bincats/USNO_A2</a>
USNO B1.0	<a href="#">I/284</a>	1,045,913,669	RAD, RECT	<a href="ftp://data.bao.ac.cn/cats/bincats/USNO_B1.0">ftp://data.bao.ac.cn/cats/bincats/USNO_B1.0</a>

表：以数据文件+查询小程序释放的星表概览

以这种方式去管理和查询星表与采用关系型数据库的方式各自有着优缺点，在此做简单对照比较。

这种方式的优点：

1. 查询程序是专门编写，是直接的文件读取，有的还采用了二级索引技术，查询执行快，这是一般的关系型数据库不能比拟的，比如象在关系型数据库中查询 USNO B1.0 这样包含 10 亿行的大星表性能已经很差。
2. 由于数据采用二进制存储，有的还采用了压缩技术，数据需要的存储空间比在关系型数据库中存储小的多。下面是一些星表采用二进制格式和采用 MySQL 数据库 MyISAM 表占用的存储容量比较。

星表名称	二进制存储格式 及大小	文本格式大小	MyISAM 格式 数据大小
UCAC1	450MB(BIN)		1.7GB
GSC-ACT	320MB(BIN)		1.0GB
UCAC2	1.6GB(BIN)		4.5GB
APM-North	11GB(BIN)	20.3GB	16.6GB

DENIS 2	14.5GB(BIN)		
GSC2.2.01	44.3GB(FITS Binary Table)		50.6GB
2MASS PSC	41.7GB(BIN)	155GB	124GB
USNO A2.0	3.5GB(PMM)		24.7GB
USNO B1.0	38.0GB(PMM)	242GB	141GB

表：星表采用二进制专用存储格式、  
无压缩文本格式和 MyISAM 表格式占用存储空间比较

### 3. 部署和使用十分简单。

采用这种方式的缺点：

1. 查询程序提供的功能十分有限，现在仅仅能提供点圈天区和矩形天区查询，以及基于 ID 号的查询。不能实现复杂的查询操作，比如统计查询等。SQL 提供了一个描述复杂查询要求的机制，这是查询程序不可能实现的。
2. 查询程序是专用的，不具备通用性，一个程序对应查询一个星表，且目前的大多数查询程序仅仅能在 Linux 和 Unix 上运行，在其他系统（Windows）上不能运行，这个原因来自于 C 语言（大部分查询程序用 C 实现）的“平台有关性”。

#### 4.2.4.2 Java 调用 C 语言程序的解决方案比较：

我们实现网格服务的编程语言是 Java（目前 GT3 仅仅支持用 Java 写的 Grid Service），但是许多天文现成软件都是 C 语言或其他语言实现的，比如我们上面提到的大星表查询程序，如果把这些程序全部用 Java 来重写，恐怕会带来巨大的工作量和长期的测试。利用这些程序的最好方法是直接调用这些程序，这样可以减少我们为实现服务所需的开发量，如何调用这些程序呢？我们探讨了三种方式：

1. 通过外部进程调用（即 Runtime.exec 方法）调用其他语言的程序，这个方法可

以适用任何语言编写的程序，也是最简单的实现方式。这种方法即首先编译部署好这些程序，然后通过Java开辟进程，调用执行本地程序命令，最后接收程序的输出，完成这些程序的调用和结果的返回。这种方式对于象Fortran语言程序这样的无别的方法与Java交互的程序来说是唯一的解决方式。但是也有一定的限制，即首先程序必须是可以用命令行格式调用的，第二是如果需要输出结果，那么输出结果是必须能被捕获和理解的，比如上述星表查询程序在屏幕输出格式标准的查询结果，我们可以用getInputStream()方法获得输出结果。还比如第四节用到的GetImage程序是将结果输出到磁盘生成文件，我们可以再次读入这个文件。如果输出方式是图形界面方式，我们就不能利用这种方式了。

通过Runtime.exec方法执行命令行程序在不同的操作系统上有不同的形式：

在Linux系统下执行命令行格式如下例所示，下例是执行ls命令。

```
String[] cmd = { "/bin/sh", "-c", "ls" };  
Process ps = Runtime.getRuntime().exec(cmds);
```

在Windows 2000或Windows NT系统下执行命令行的例子如下，是执行dir命令。

```
Process p=Runtime.getRuntime().exec("cmd /c dir");
```

在Unix系统下执行命令行的例子如下，是执行date命令。

```
Process p=Runtime.getRuntime().exec("/usr/bin/sh -c date");
```

采用这种方式的弊端是需要开辟进程和捕获程序输出，并关闭进程，比较耗时，系统开销大。

2. Java 本机接口 (Java Native Interface)，或者称为 JNI，提供了Java调用C，C++编写的动态链接库的机制。实现这种方式需要编写生成和部署动态链接库（在WINDOWS平台上是DLL文件形式，在UNIX机器上是SO文件形式），和用Java编写调用链接库的类，开发量比较前一种方式要大些。而且当本地链接库运行时，没有有效地防数组越界错误、错误指针引用带来的间接错误等。所以必须保证本地代码的稳定性，因为，丝毫的错误都可能导致Java虚拟机崩溃
3. 可以通过TCP/IP实现Java代码与本地其他程序代码的交互工作，比如MySQL数据库与Java的交互就是采用TCP/IP端口通讯的方式进行交互。这个开发量更大，



需要去编写实现通讯协议方式，阻碍了广泛应用。

#### 4.2.4.3 基于文件系统的星表Cone-Search服务的实现

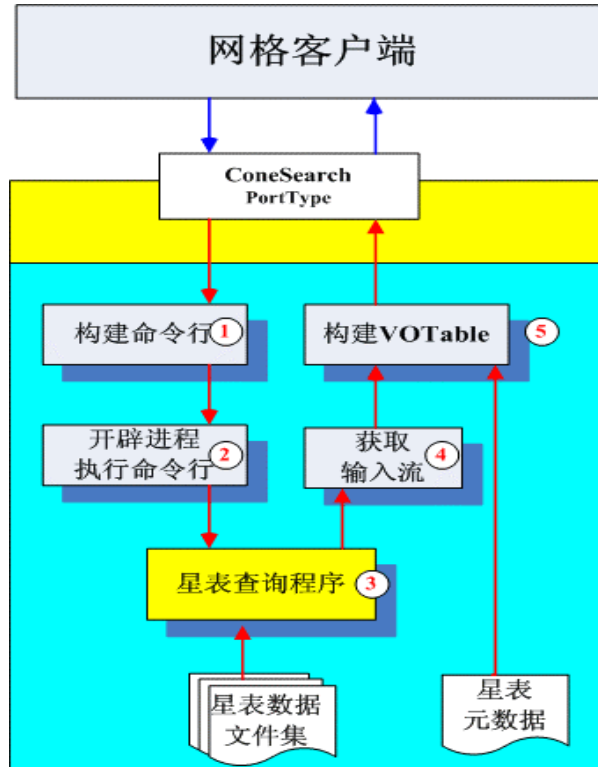


图4. ：基于文件系统的星表Cone-Search服务实现流程图

1. 将 ConeSearch 接口传来的查询参数转换成查询程序的命令行；如果查询参数不合法，抛出 `ErrorInputParametersException` 异常。比如查询 GSC1.2 星表，输入参数  $ra=30.0$ ,  $dec=20.0$ ,  $radius=0.20$ ，转换成的命令行为：

```
/data2/GSC1.2/gsc -c 02 00 00.0 +20 00 00.0 -r 12.0 -n 1000
```

- `/data2/GSC1.2/gsc`：可执行程序的路径和名称；
- `-c 02 00 00.0 +20 00 00.0`：指定查询天区中心点；
- `-r 12.0`：指定查询为点圈天区查询，查询半径为 12 角分（0.20 度）；
- `-n 1000`：指定最大返回行数为 1000 行；

查询其他星表的命令行与上例基本一致。

2. 通过开辟外部进程调用（即 Java 的 `Runtime.exec` 方法），执行命令行，我们采用了这种最简单的 Java 调用 C 程序的办法。
3. 星表查询程序读取数据文件，导出查询结果数据。
4. “捕获”查询程序返回的查询结果。
5. 读取星表元数据和列（数据字段）元数据信息，与查询结果构造生成 `VOTable`，返回给网格客户端。在这里我们采用了 `StringBuffer` 来建立 `VOTable`，由于 `VOTable` 的星表元数据和列元数据不会因为查询参数的变化而变化，只有 `PARAM` 和 `TABLEDATA` 的内容会变化，所以 `VOTable` 元数据部分是预先在 `StringBuffer` 构造好的，节省了构造 `VOTable` 所需要的开销。

### § 4.3 星表查询服务

## § 4.4 简单天文图像获取服务的设计和实现

天文图像数据占天文数据的相当大的比例，天文图像的数据模式和格式多种多样，天文图像数据库或程序提供的查询功能也是各有不同，如何制定一个标准的天文图像查询网格服务标准，将天文图像数据融入到网格中去，我们基于目前天文学家的查询一般要求和目前天文图像数据库一般所能提供的查询功能，设计了一个最基本的简单天文图像查询网格服务接口，并基于目前拥有的天文图像数据库实现了 DSS I 天文图像获取服务。

### 4.4.1 简单图像获取服务接口

简单图像获取服务由接口 `AccessSimpleImagePortType` 定义，其特定的方法只有一个，描述如下：

- `getImage`

1. 方法原型：`byte[] getImage(double ra,double dec,double width,double height)`

2. 功能：类似于第二节提到的矩形天区搜索（Rectangle Search），即查询的天区图像为矩形，且图像第二轴与赤道坐标系的经线夹角为 0 度，是一个没有旋转的图像。这种方法是绝大多数天文图像数据库或查询程序能提供的查询方式，也是其中的大部分唯一能提供的查询方式。

3. 参数：

参数名称	数据类型	参数意义	单位
ra	double	查询矩形天区中心点赤经, J2000 历元, Fk5 坐标系统	度
dec	double	查询矩形天区中心点赤纬, J2000 历元, J2000 历元, 坐标系统 Fk5	度
width	double	矩形天区赤经方向的相对于天球球心的张角	度
height	double	矩形天区赤纬方向的相对于天球球心	度

		的张角	
--	--	-----	--

表: getImage 方法的输入参数

需要注意的是: width 是在赤经方向上相对于天球球心的张角不是沿赤纬线跨的经度, 经度跨度=张角/COS (dec)。

4. **返回值:** 返回一图像的字节数组。这个图像的格式为采用 Gzip 方式压缩的 FITS 图像。具体规范说明如下:

**图像格式:** 服务返回的查询结果的必须是单个图像, 图像必须为 FITS 格式, 应该包含一个有效的 FITS WCS (世界坐标系) 信息, 而且不允许包含无效的象元数据, 如果需求的区域没有数据值, 那就用一个值代替空值, 并在 FITS 用 BLANK 关键字表明这个值。

**图像压缩方法:** 因为 FITS 格式文件比较大, 所以采用最通用的 Gzip 方式进行压缩后返回图像。

失败—抛出异常, 若输入参数非法, 抛出 `ErrorInputParametersException` 异常, 出现其他错误抛出 `ServiceInternalErrorException`。

注: 抛出 `ErrorInputParametersException` 条件: 当出现输入参数不在以下合法范围内时

- A)  $0.0 \leq ra < 360.0$ ;
- B)  $-90.0 \leq dec \leq +90.0$ ;
- C)  $0.0 \leq width \leq 1.0$ ;
- D)  $0.0 \leq height \leq 1.0$ ;

#### 4.4.2 DSS-I 天图获取服务的实现:

DSS-I 天图是天文界常用到的天图, 具体介绍如下, 我们首先实现了实现了 DSS-I 天图的简单图像获取服务, 服务的架构比较简单, 因为有实现查询导出 DSS-I 图像的 `GetImage` 程序, 我们需要做的是将这个程序封装成服务, 来实现我们需要的服务功能。

#### 4.4.2.1 DSS-I 天图简介:

DSS 天图是采用 Palomar 和 UK 斯密特望远镜对全天进行底片巡天,并由 STScI 数字化得到的, DSS-I 是第一期巡天的结果。每一张底片覆盖  $6.5 \times 6.5$  度天区,采用改进的 PDS 显微光度计扫描数字化得到的,每个象素的大小为 25 微米 (1.7 角秒每像素) 或 15 微米 (1 角秒每像素),得到的数字图像大小为  $14000 \times 14000$  或  $23040 \times 23040$  像素。为了方便数据的释放,这些数字图像采用 H-compress 方式压缩,典型的压缩率为 10,每张数字底片分成 640 个文件存储, DSS-I 共采用 100 张 CDROM 释放,数据量 60GB,这些数据可以通过 GetImage 程序或 GetDSS 程序查询导出。

#### 4.4.2.2 GetImage 程序介绍:

在这个服务的实现中我们主要使用了一个导出 DSS-I 压缩图像数据的程序 GetImage, GetImage 采用 C 语言编写,只能在 Linux 或 Unix 下运行,编译设置完成后,调用时给出坐标中心点和天区大小, GetImage 将会读出相关 DSS-I 数据,解压拼接后生成图像文件,输入方式可以是交互式的也可以从文件读入。

一个简单的生成 Fits 图像文件的命令如下:

```
echo "sang 10 28 32.5 +20 30 40.9 28 40.5" |getimage >&/dev/null
```

主要参数如下:

- `getimage`: 为命令的名称;
- 10 28 32.5 为查询天区中心点的赤经,时 分 秒格式,默认历元为 J2000,坐标系统为 Fk5;
- +20 30 40.9 查询天区中心点的赤纬,度 角分 角秒格式,在度前正负号表示北南纬,默认历元为 J2000,坐标系统为 Fk5;
- 28 为图像在赤经方向上的张角,单位:角分;
- 40.5 为图像在赤纬方向上的张角,单位:角分;

- sang 为生成的图像文件的头四个字母，GetImage 程序会生成一个以 sang 后面再加上 4 位数字表示的天区序号为文件名的 FITS 文件。如 sang0401.fits。

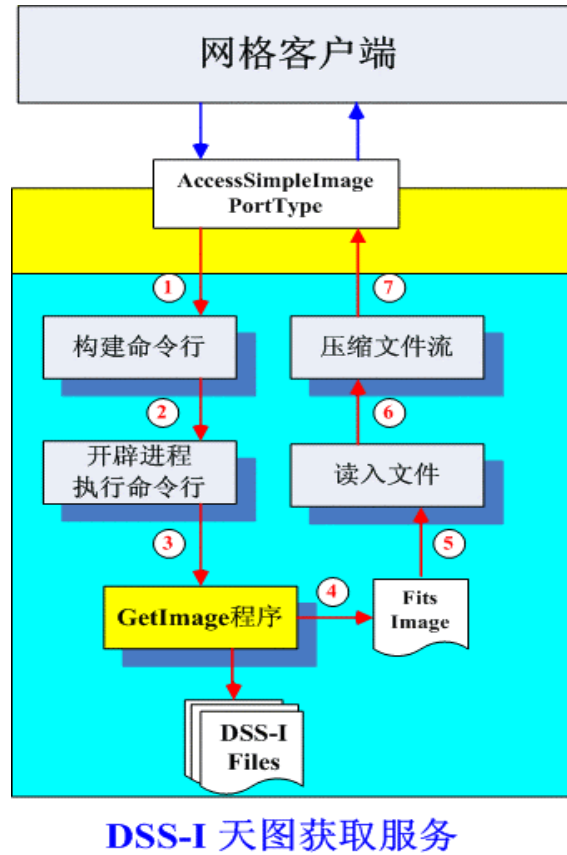


图 4. : DSS-I 天图简单图像获取服务实现流程图

#### 4.4.2.3 DSS-I 天图简单图像获取服务实现流程

DSS-I 天图简单图像获取服务实现流程如图 4.所示，总共分为五步，具体说明如下：

1. 接口传来的查询参数转换成调用 GetImage 程序的命令行；如果查询参数不合法，抛出 `ErrorInputParametersException` 异常。

为了避免生成的 Fits 文件名称出现重复而导致冲突(这出现的几率很小，但是也要避免)，所以我们为要输出的 fits 图像文件设定一个 10 个字符长度的随机文件名，如 6cOx3L212r。

比如输入参数 `ra=120.0, dec=30.0, width=0.5, height=0.4`，那么生成的

GetImage 命令行为:

```
echo "6cOx3L212r 05 00 00.0 +30 00 00.0 30.0 24.0" |getimage >&/dev/null
```

2. 通过开辟外部进程调用（即 Java 的 `Runtime.exec` 方法），执行命令行。
3. `GetImage` 程序读出相关文件数据，解压拼接后生成 fits 图像文件；
4. 找到并将 fits 图像文件读入文件流，读入后删除这个 fits 文件；

由于 `GetImage` 生成的文件的名称等于我们设定的随机文件名，加上表示天区的 4 位序号，这四位序号我们不得而知，我们必须先按文件前边的随机文件名“找到”这个文件，再读入。

5. 采用 `Gzip` 压缩方式（使用 `java.util.zip` 类）压缩 fits 图像文件流，并转换成字节数组通过服务接口返回给网格客户端。

## 参考文献

- [1] 崔辰州. 博士学位论文 中国虚拟天文台系统设计. 2003(40-47)
- [2] Borja Sotomayor. 《The Globus Toolkit 3 Programmer's Tutorial》  
<http://www.casa-sotomayor.net/gt3-tutorial/>
- [3] 李彬. 用GT3 开发网格服务  
<http://www-900.ibm.com/developerWorks/cn/webservices/ws-gt3/index.shtml>



## 第五章 数据可视化服务的实现

### § 5.1 VOTable XSLT 样式表的设计

#### 5.1.1 XSLT<sup>[1]</sup>简介:

XSLT是XML大家族的成员之一，是eXtensible Stylesheet Language: Transformation的缩写，即可扩展样式语言：转换。XSLT被设计用来提供一套规则，用于将一个XML文档转换为另外一个文档，这个文档可以是XML格式，HTML格式，甚至是自定义的文本格式，比如PDF，CSV格式。XSLT的一个更通用全面的定义是——XSLT是一种用来变换XML结构的语言<sup>[2,3,4,5,6]</sup>。

为何要对XML进行转换呢？

(1) 为了方便人们阅读，我们知道，XML语言是一个分离数据与显示的语言，它并不便于人们的阅读理解，所以要将其转换为人们容易阅读和打印的文档，比如最常见的HTML，PDF甚至是一段声音等等，XSLT就提供了转换的一种机制和描述方法<sup>[7,8]</sup>。目前XSLT最常用的就是将XML转换为HTML格式文档，其目的是可以在网络浏览器（如IE）中进行显示，它克服了网络浏览器在支持XML方面整体上缺乏连续性引起的两难问题。HTML有着比XML更好的向后兼容性，因此可以在XML中描述和存储数据，而用HTML显示数据。这也就意味着各种用户，包括那些使用老版本浏览器的用户，都可以通过他们的浏览器看到此HTML文档<sup>[9,10]</sup>。

(2) 为了在不同的应用程序之间传递数据。我们需要能把数据从一个应用程序所使用的数据模型变换成另一个应用程序所使用的数据模型。为了把这些数据装入到一个应用程序中去，必须的格式可能是CSV格式文件，一个SQL脚本等，虚拟天文台中的一个例子是，描述天文星表数据查询要求的标准语言是XML格式的ADQL语言，但各关系数据库的标准查询语言是SQL，在执行数据库查询时必须将ADQL转换成SQL语句。该格式也可能是另外一个XML文档，但该文档采用了不同的词汇和结构。随着XML的普及，XSLT在应用程序间数据转换中的作用也会更加重要<sup>[11]</sup>。

XSLT规范没有为转换输出格式定义要求，但提供了推荐的输出类型<sup>[12]</sup>：

(1) XML: 输出文档可以是XML文档或XML片段，XML是默认的输出方式；

(2) HTML: 输出文档符合HTML 4.0<sup>[13]</sup>规范, 若想产生XHTML, 应该使用XML输出方式。

(3) 文本: 输出文档可以是任何一种基于文本的方式, 比如 CSV, PDF 等。

XSLT是表达一种格式良好的XML文档的语言, 这种文档称为“样式表(Stylesheet)”, XSLT 样式表也是一个 XML 文档, 通过使用 XML 的尖括号标记语法来表示文档的结构。它是一个声明将XML转换为其他格式规则的文件, 包含一组作为转换规则和指令的元素。样式表的语法规则非常复杂, 在此不做介绍, 详细参见<sup>[14]</sup>。

XSLT 通常是将 XML 数据从一种形式转换成另一种形式的最方便的方法, 需要的编码工作量要少于使用低级 DOM 和 SAX 接口编码的过程性应用程序进行转换。XSLT 为 XML 的转换提供了很好的机制, 即采用编写好的 XSLT 样式表对 XML 进行灵活转换, 而且如果需要对生成的格式进行变动时, 仅仅需要对样式表进行修改, 而没有必要去修改程序。

使用 XSLT 样式表和需要转换的 XML 文档并不能自动执行变换, 还必须有一个 XSLT 处理器(XSLT Processor), XSLT 处理器是一个应用模块, 它负责将 XSLT 样式表规定的转换规则应用到需要转换的 XML 文档, 并生成结果文档。XSLT 转换机制如下图所示:

目前有许多可用的XSLT处理器, 他们支持不同类型的转换输出格式。大多数都支持把XML根据XSLT转换成HTML或纯文本的能力。目前常用到的XSLT处理器有Xalan<sup>[15]</sup>、Saxon<sup>[16]</sup>和微软的MSXML。最近推出的浏览器, 比如Internet Explorer 5.5及更高版本、Netscape 6.1及更高版本和Mozilla等, 都内置了XSLT处理器, 它们都支持XSLT转换。

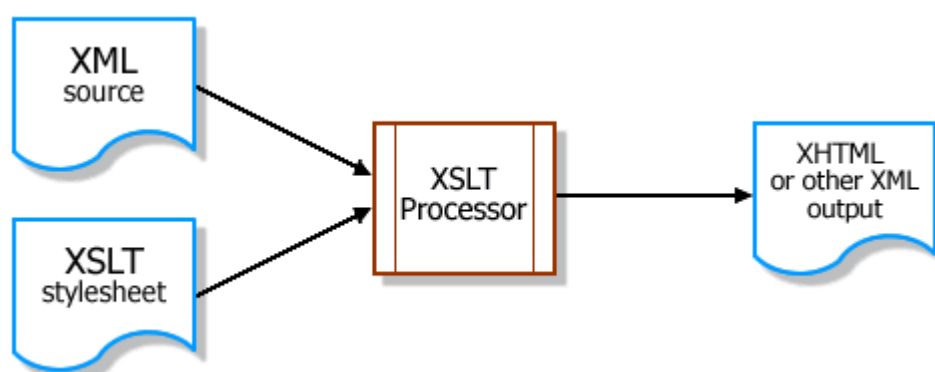


图 5.1: XSLT处理器利用XSLT样式表将XML转换成其他格式文档<sup>[17]</sup>

### 5.1.2 用于将 VOTable 格式转换成 CSV 格式的样式表

VOTable是虚拟天文台推出的用来封装天文表列数据的XML格式标准，它包括了非常复杂的元素结构，用来包含数据表格和元数据信息，在浏览器中显示其复杂的树型结构会非常的冗长，很不方便用户的阅读，用户一般仅仅需要真正的表格数据信息加上列的描述，而忽略掉其他辅助信息，所以我们设计了一个简洁的样式表，将VOTable中的表格数据加上相应的列名转换成的CSV（逗号隔开每个数据值）格式，其他的辅助信息将不做转换。CSV格式比较方便用户的阅读，它在VOTable格式推出以前是天文星表的交换格式之一，更重要的是CSV格式很容易被导入Excel或某些关系数据库。CSV本身就是Excel表的存储格式，也是大部分关系数据库（MySQL<sup>[18]</sup>，PostgreSQL<sup>[19]</sup>等）能直接导入的数据格式。

目标 CSV 文档具体格式如下：CSV 格式文档的第一行以“#”开头，后面是资源的名称（RESOURCE 元素的 name 属性值），第二行是列的名称（FIELD 元素的 name 属性值），每个列名以“#”开头，列名间用逗号隔开，后面的行是表格数据（TD 元素的内容）。每一行都以换行符结束，每行的每一个数据项间用逗号隔开。

在此不对样式表的结构做详细分析，关于具体样式表的语法参见，在此仅说明几点：

(1) XSLT1.0 的名称空间为<http://www.w3.org/1999/XSL/Transform>，样式表的最外层元素必须为 `xsl:stylesheet`，所以第二行（`<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">`）表示本文档是一个合法的XSLT样式表。

(2) 第三行（`<xsl:output method="text">`）是设定文档的输出格式为文本格式。

(3) 文档中加粗的部分，是在生成的 CSV 文档中每一行结束时加换行符（LF）。

(4) 表的各行的缩进是为了显示而设定，在真实的文档中各行没有任何用空格或 TAB 键组成的缩进，各行之间也没有回车键和换行键，因为这些将会在转换中被作为合法字符，从而影响 CSV 文档的格式。

表 5.1 是需要转换的 VOTable 文档，而表 5.2 是用样式表（见表 5.3）转换成的 CSV 格式文档。

---

```
<?xml version="1.0"?>
<!DOCTYPE VOTABLE SYSTEM "http://www.lamost.org/~sang/VOTable.dtd">
<VOTABLE version="1.0">
  <DEFINITIONS><COOSYS ID="myJ2000" equinox="2000." epoch="2000" system="eq_FK5"/></DEFINITIONS>
  <RESOURCE name="GSC1.2">
    <DESCRIPTION>
      This is an excerpt of the HST Guide Star Catalog, Version 1.2 (Lasker+ 1996).</DESCRIPTION>
    <TABLE>
```

---

---

```

<FIELD ID="ra" name="RA(J2000)" ref="myJ2000" ucd="POS_EQ_RA_MAIN" unit="deg"
datatype="double" precision="F5">
  <DESCRIPTION>Right ascension in J2000, epoch of plate</DESCRIPTION>
</FIELD>
<FIELD ID="dec" name="Dec(J2000)" ref="myJ2000" ucd="POS_EQ_DEC_MAIN" unit="deg"
datatype="double" precision="F5">
  <DESCRIPTION>Declination in J2000, epoch of plate</DESCRIPTION>
</FIELD>
<FIELD ID="mag" name="Pmag" ucd="PHOT_PHG_MAG" unit="mag" datatype="float" width="5"
precision="2">
  <DESCRIPTION>photographic magnitude (see n_Pmag)</DESCRIPTION>
</FIELD>
<LINK content-role="doc" title="doc" href="http://vizier.u-strasbg.fr/viz-bin/Cat?I/254"/>
<DATA>
  <TABLEDATA>
    <TR><TD>4.7766</TD><TD>72.8474</TD><TD>8.59</TD></TR>
    <TR><TD>5.4576</TD><TD>72.6528</TD><TD>12.18</TD></TR>
    <TR><TD>3.9867</TD><TD>72.9484</TD><TD>12.09</TD></TR>
    <TR><TD>8.9587</TD><TD>72.6635</TD><TD>14.38</TD></TR>
    <TR><TD>5.4847</TD><TD>72.8272</TD><TD>14.96</TD></TR>
  </TABLEDATA>
</DATA>
</TABLE>
</RESOURCE>
</VOTABLE>

```

---

表 5.1: 被转换的 VOTable 文档

---

```

#Resource Name:GSC1.2
#RA(J2000),#Dec(J2000),#Pmag
4.7766,72.8474,8.59
5.4576,72.6528,12.18
3.9867,72.9484,12.09
8.9587,72.6635,14.38
5.4847,72.8272,14.96

```

---

表 5.2: 转换后得到的 CSV 格式文档

---

```

<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:output method="text" omit-xml-declaration="yes" indent="no"/>

```

---

---

```

<xsl:template match="/">
  <xsl:for-each select="VOTABLE">
    <xsl:for-each select="RESOURCE">#Resource Name:<xsl:value-of select="@name"/>
      <xsl:value-of select="'&#xA;'" />
      <xsl:for-each select="TABLE">
        <xsl:for-each select="FIELD">#<xsl:value-of select="@name"/>
          <xsl:if test="position() != last()">
            <xsl:value-of select="','" />
          </xsl:if>
        </xsl:for-each>
        <xsl:value-of select="'&#xA;'" />
        <xsl:for-each select="DATA">
          <xsl:for-each select="TABLEDATA">
            <xsl:for-each select="TR">
              <xsl:for-each select="TD">
                <xsl:value-of select=".'" />
                <xsl:if test="position() != last()">
                  <xsl:value-of select="','" />
                </xsl:if>
              </xsl:for-each>
            </xsl:for-each>
            <xsl:value-of select="'&#xA;'" />
          </xsl:for-each>
        </xsl:for-each>
      </xsl:for-each>
    </xsl:for-each>
  </xsl:template>
</xsl:stylesheet>

```

---

表 5.3: 用于将 VOTable 格式转换成 CSV 格式的样式表

### 5.1.3 用于将 VOTable 格式转换成 HTML 格式的样式表

最方便用户阅读也是最常用的格式是HTML，所以最好将VOTable文档转换成HTML格式显示，美国虚拟天文台（NVO）给出了一个将VOTable转换为HTML的XSLT样式表(<http://us-vo.org/xml/VOTable-basic.xsl>)，但转换成的HTML美观性上和完整性上不尽如人意（如图所示），本人根据需求重新编写了一个XSLT样式表，可以对VOTable 1.0 版进行良好的转换，其转换生成的文档符合HTML 4.0 标准，显示

更加美观，信息也更加丰富，这个样式表已经被应用到实现我们的星表可视化显示服务中去。

这个样式表将 VOTable 文档中大部分重要的信息都转换到 HTML 文档中，其中包括：

- (1) 资源的名称 (RESOURCE 元素的 name 属性值)；
- (2) 资源的描述 (RESOURCE 元素的子元素 DESCRIPTION 的内容)；
- (3) 参数的所有信息 (所有 PARAM 元素的所有属性值)；
- (4) 有关列的所有元数据信息 (所有 FIELD 元素的所有属性值)；
- (5) 表格数据 (即所有 TD 元素的内容)。

目前可以通过网址<http://www.lamost.org/~sang/VOTable-ChinaVO.xsl>获取到这个样式表文档，如果通过浏览器 (如IE 6.0) 转换显示VOTable格式文档时，可以在文档中引用这个样式表。具体加入VOTable文档的语句如下：

```
<?xml-stylesheet type="text/xsl" href="http://www.lamost.org/~sang/VOTable-ChinaVO.xsl"?>
```

在用样式表转换后生成的HTML还引用到一个层叠样式表 (CSS)，指向如下地址：<http://www.lamost.org/~sang/rlamost.css>，采用这个独立的层叠样式表文件，来代替在XSLT样式表中嵌入大量关于字体、颜色和其它的样式元素。XSLT转换产生的HTML只是指向这个独立的CSS文件，这就使XHTML更小，同时也简化了XSLT样式表文档，也使得浏览器下载页面时速度更快。这样XSLT和CSS分别执行不同的任务并相互补充。XSLT负责把XML转换为其他的格式，如HTML或WML，而CSS只是定义表达的样式。

图 5.2 和图 5.3 显示的分别为采用 NVO 和我们提供的样式表转换得到的 HTML。

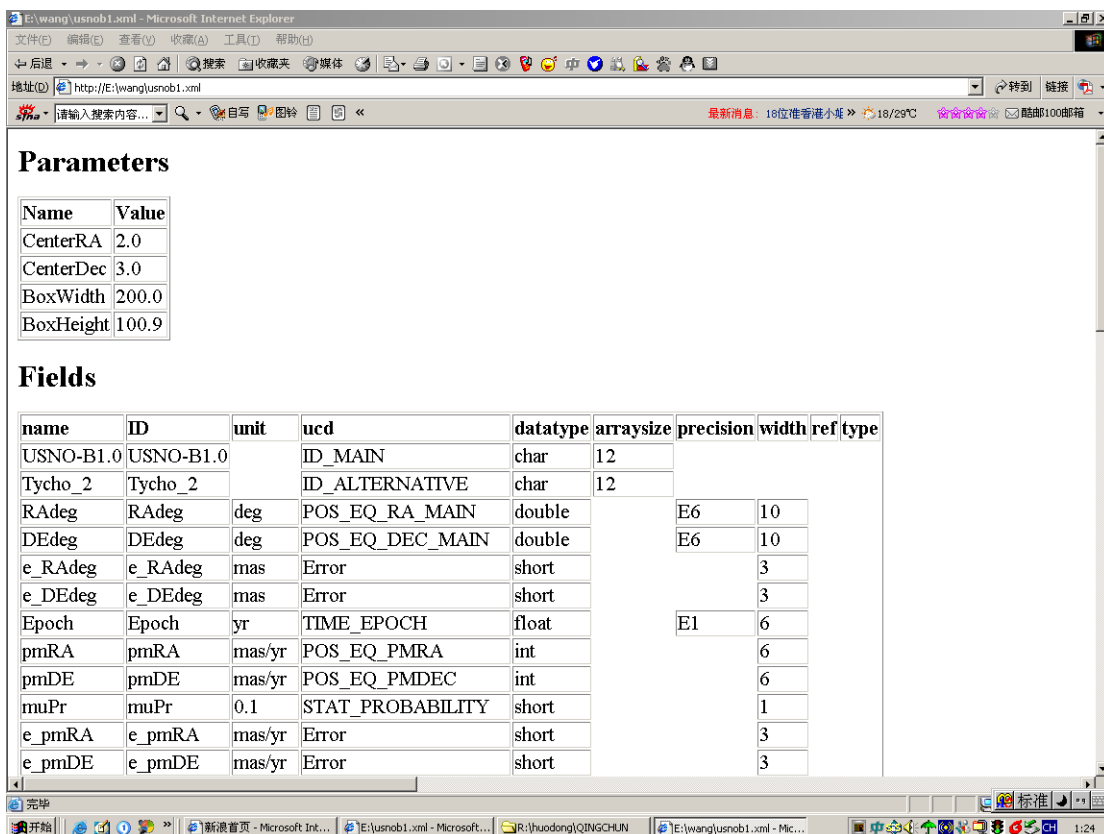


图 5.2: 采用 NVO 提供的样式表转换得到的 HTML 显示

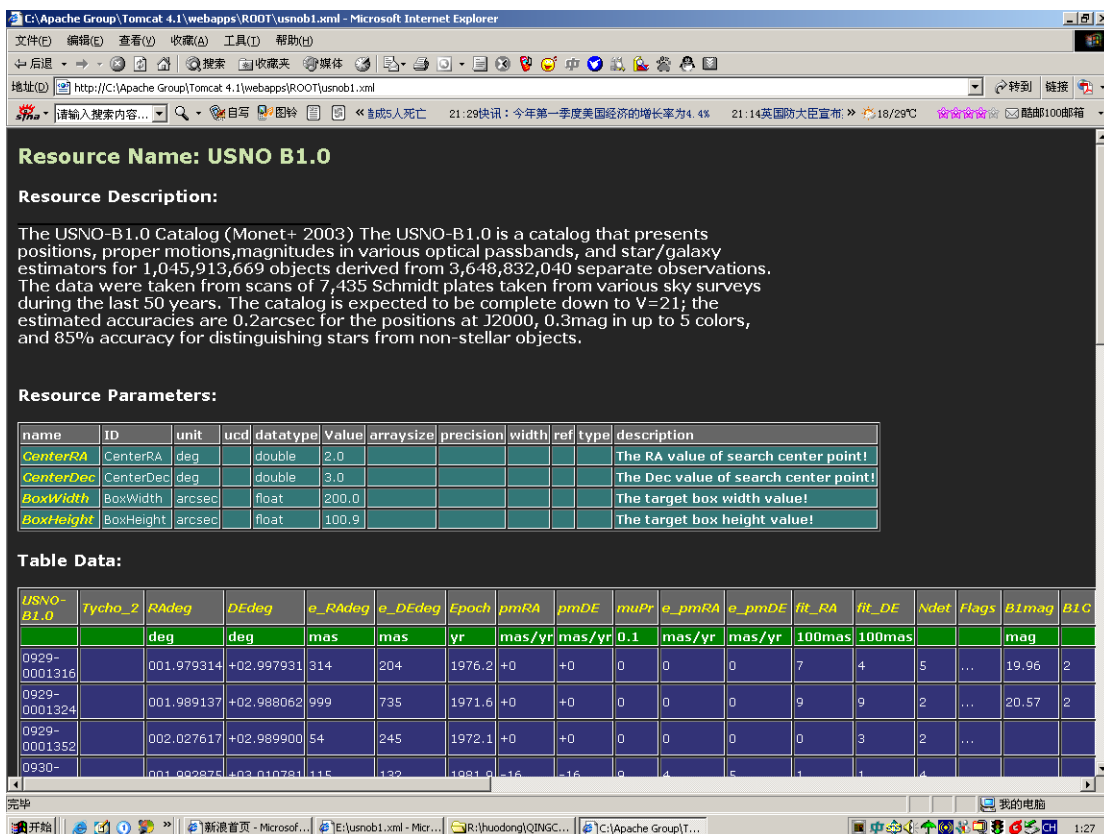


图 5.3: 采用我们编写的样式表转换得到的 HTML 显示



```

<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
                xmlns:HTML="http://www.w3.org/Profiles/XHTML-transitional">
  <xsl:output method="html"/>
  <xsl:template match="/">
    <link rel="stylesheet" type="text/css" href="http://www.lamost.org/~sang/rlamost.css"
    title="Style"/>
    <HTML>
      <HEAD/>
      <BODY>
        <table bgcolor="#000000">
          <xsl:for-each select="VOTABLE">
            <xsl:for-each select="RESOURCE">
              <H1>Resource Name: <xsl:value-of select="@name"/></H1>
              <H2>Resource Description:</H2>
              <table width="800">
                <H6><font size="0" color="white"><xsl:value-of select="DESCRIPTION"/></font></H6>
              </table>

              <H2>Resource Parameters:</H2>
              <table border="1" frame="box">
                <tr>
                  <td bgcolor="#666666"><en><b>name</b></en></td>
                  <td bgcolor="#666666"><en><b>ID</b></en></td>
                  <td bgcolor="#666666"><en><b>unit</b></en></td>
                  <td bgcolor="#666666"><en><b>ucd</b></en></td>
                  <td bgcolor="#666666"><en><b>datatype</b></en></td>
                  <td bgcolor="#666666"><en><b>Value</b></en></td>
                  <td bgcolor="#666666"><en><b>arraysize</b></en></td>
                  <td bgcolor="#666666"><en><b>precision</b></en></td>
                  <td bgcolor="#666666"><en><b>width</b></en></td>
                  <td bgcolor="#666666"><en><b>ref</b></en></td>
                  <td bgcolor="#666666"><en><b>type</b></en></td>
                  <td bgcolor="#666666"><en><b>description</b></en></td>
                </tr>
                <xsl:for-each select="PARAM">
                  <tr>
                    <td bgcolor="#337777"><em><xsl:value-of select="@name"/></em></td>
                    <td bgcolor="#337777"> <xsl:value-of select="@ID"/> </td>
                    <td bgcolor="#337777"> <xsl:value-of select="@unit"/> </td>
                    <td bgcolor="#337777"><font color="red"><xsl:value-of select="@ucd"/></font></td>
                    <td bgcolor="#337777"> <xsl:value-of select="@datatype"/> </td>
                    <td bgcolor="#337777"> <xsl:value-of select="@value"/> </td>

```

```

        <td bgcolor="#337777"> <xsl:value-of select="@precision"/> </td>
        <td bgcolor="#337777"> <xsl:value-of select="@arraysize"/> </td>
        <td bgcolor="#337777"> <xsl:value-of select="@width"/> </td>
        <td bgcolor="#337777"> <xsl:value-of select="@ref"/> </td>
        <td bgcolor="#337777"> <xsl:value-of select="@type"/> </td>
        <td bgcolor="#337777"><b><xsl:value-of select="DESCRIPTION"/></b></td>
    </tr>
</xsl:for-each>
</table>

<xsl:for-each select="TABLE">
    <H2>Table Data:</H2>
    <table border="1" frame="box">
        <tr>
            <xsl:for-each select="FIELD">
                <td bgcolor="#666666"><em><b><xsl:value-of select="@name"/></b></em></td>
            </xsl:for-each>
        </tr>
        <tr>
            <xsl:for-each select="FIELD">
                <td bgcolor="Green"><b><xsl:value-of select="@unit"/></b></td>
            </xsl:for-each>
        </tr>
        <xsl:for-each select="DATA">
            <xsl:for-each select="TABLEDATA">
                <xsl:for-each select="TR">
                    <tr>
                        <xsl:for-each select="TD">
                            <td bgcolor="#333377"><xsl:value-of select="."/></td>
                        </xsl:for-each>
                    </tr>
                </xsl:for-each>
            </xsl:for-each>
        </xsl:for-each>
    </table>
    <H2>Table Fields:</H2>
    <table width="90%" border="1" frame="box">
        <tr>
            <td bgcolor="#666666"><en><b>name</b></en></td>
            <td bgcolor="#666666"><en><b>ID</b></en></td>
            <td bgcolor="#666666"><en><b>unit</b></en></td>
            <td bgcolor="#666666"><en><b>ucd</b></en></td>
            <td bgcolor="#666666"><en><b>datatype</b></en></td>
            <td bgcolor="#666666"><en><b>arraysize</b></en></td>

```

---

```

        <td bgcolor="#666666"><en><b>precision</b></en></td>
        <td bgcolor="#666666"><en><b>width</b></en></td>
        <td bgcolor="#666666"><en><b>ref</b></en></td>
        <td bgcolor="#666666"><en><b>type</b></en></td>
        <td bgcolor="#666666"><en><b>description</b></en></td>
    </tr>
    <xsl:for-each select="FIELD">
    <tr>
        <td bgcolor="#337777"><em><xsl:value-of select="@name"/></em></td>
        <td bgcolor="#337777"> <xsl:value-of select="@ID"/> </td>
        <td bgcolor="#337777"> <xsl:value-of select="@unit"/> </td>
        <td bgcolor="#337777"><font color="red"><xsl:value-of
select="@ucd"/></font></td>
        <td bgcolor="#337777"> <xsl:value-of select="@datatype"/> </td>
        <td bgcolor="#337777"> <xsl:value-of select="@arraysize"/> </td>
        <td bgcolor="#337777"> <xsl:value-of select="@precision"/> </td>
        <td bgcolor="#337777"> <xsl:value-of select="@width"/> </td>
        <td bgcolor="#337777"> <xsl:value-of select="@ref"/> </td>
        <td bgcolor="#337777"> <xsl:value-of select="@type"/> </td>
        <td bgcolor="#337777"><b><xsl:value-of select="DESCRIPTION"/></b></td>
    </tr>
    </xsl:for-each>
    </table>
    </xsl:for-each>
    </xsl:for-each>
    </xsl:for-each>
    </xsl:for-each>
    </table>
</BODY>
</HTML>
</xsl:template>
</xsl:stylesheet>

```

---

表 5.4: 用于将 VOTable 格式转换成 HTML 格式的样式表

## § 5.2 Web 环境下 VOTable 数据显示的几种解决方案

VOTable 格式现在已经成为天文表列数据特别是星表的标准交换格式，无论是我们实现的网格环境下的星表获取服务，还是其他虚拟天文台星表查询 Web 服务，都是采用 VOTable 来封装星表的数据和元数据信息，VOTable 格式已经成为大势所趋。但是一个突出的问题是 VOTable 格式数据 VOTable 符合 XML 格式标准，虽然方便了计算机程序之间的数据交换，但是 XML 是将数据与显示分离的文档，所以 VOTable 并不适合人来阅读，如何基于目前 Web 服务器/用户浏览器服务架构更好的在用户浏览器上显示 VOTable 数据，是我们需要研究的问题。

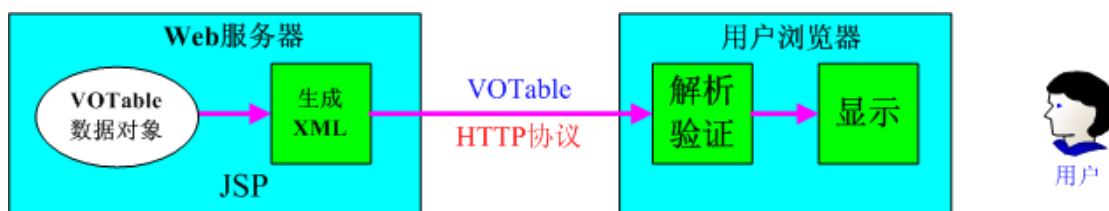
我们在 VOTable 数据的可视化方面做了一些有益的尝试，我们基于现有的 Web Server/Client 服务结构设计实现了几种可视化 VOTable 数据显示界面。我们按照 VOTable 的显示方式将可视化方法分为三种：

1. VOTable 格式直接显示。VOTable 符合 XML 格式，所以能直接发送到 Web 浏览器显示，默认的显示方式按树型结构显示所有内容，这不方便用户的阅读。
2. 将 VOTable 格式转换成 HTML 格式显示。我们将 VOTable 转换成适合显示的 HTML，并将其中最有用的信息以表格的形式呈现给用户，提高了数据的可阅读性，这一种是最常用的方式，我们的技术实现方式也最多。
3. 将 VOTable 格式转换成图形显示。采用图形显示 VOTable 数据是最美观的可视化方法。

我们采用不同的技术解决手段实现了上述三种可视化 VOTable 数据的 Web 界面。对于将 VOTable 格式转换为 HTML 格式显示，我们尝试采用了三种不同的技术途径来实现，并对其优缺点进行了对照比较。对于将 VOTable 格式转换为图形格式显示，我们实现了一种方案，并提出了另外一种解决方案的构想。

### 5.2.1 设计实现直接在用户浏览器显示 VOTable 数据

这种方法即是把 Web 服务器端生成的 VOTable 文档直接发送给用户的浏览器。如图所示，



在服务器端，首先生成 VOTable 格式对象，这个对象可能是调用我们上一章提到的星表获取服务得到的，或者是基于本地的数据库生成的。这个生成 VOTable 数据对象用 JSP 生成 XML 格式页面，然后将页面通过 HTTP 协议返回给用户的浏览器上。生成 XML 格式页面的主要要求是恰当的设置页面的内容类型。需要如下语句指定页面为 XML 类型页面。

```
<% @ page contentType="text/xml" %>
```

现在的浏览器如 IE, Mozilla, Netscape 6.1 等都内嵌了对 XML 进行验证的解析器，能对 VOTable 格式文档进行格式有效性验证（根据 VOTable 文档指向的 DTD 或 XSD），如果格式正确将按照默认的树型格式显示整个文档。

这种方式最简单也最容易实现，用户也可以不使用浏览器而是编写网络程序接收解析 VOTable 文档，所以这种方式适合计算机程序之间的交互，通用性好。但是其不适合用户通过浏览器去阅读，原因如下：第一是某些老的浏览器不支持 XML 格式文档的显示，第二是 XML 格式文档结构和标志非常复杂，其默认的树型结构显示并不方便阅读。如图 5.4 所示：

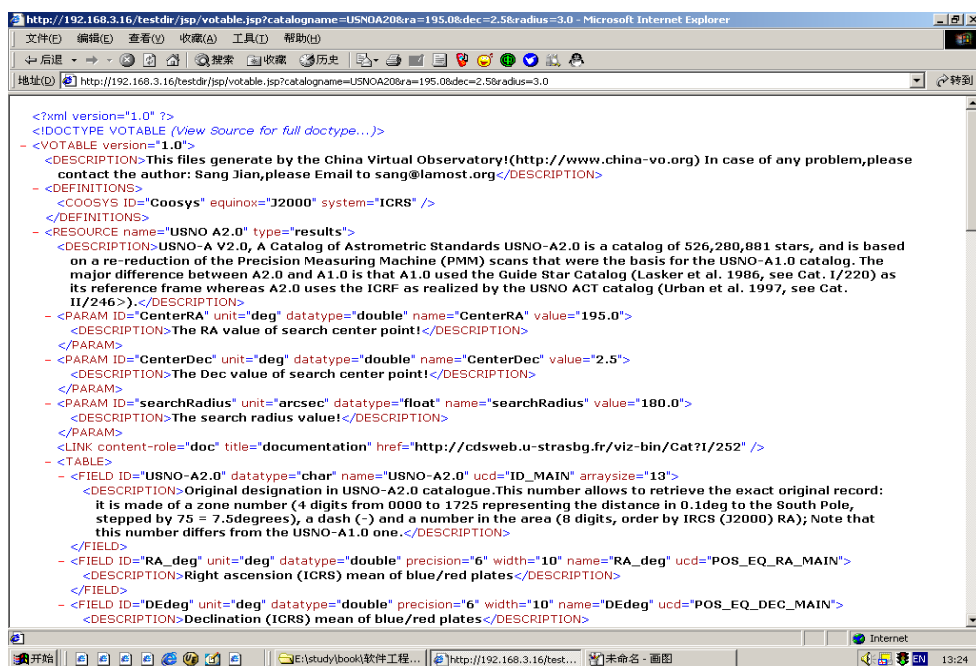


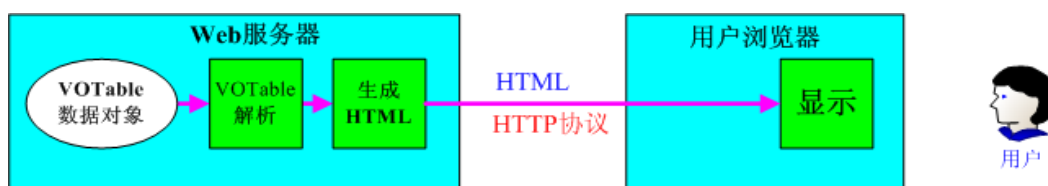
图 5.4: VOTable 格式页面的直接显示

## 5.2.2 设计实现以 HTML 格式显示 VOTable 数据的界面

我们知道 HTML 用来描述数据的外观，将 VOTable 数据以 HTML 的格式显示是最直观，也是最方便用户阅读浏览，而在将 VOTable 转换成 HTML 显示时可以显示用户需要浏览的信息，而省略掉不重要的辅助信息（比如 LINK 元素的内容等）。

将 VOTable 转换成 HTML 格式有多种实现方式，我们实现了三种转换的方式：

1. 在服务器端解析 VOTable 格式数据，然后转换为 HTML 文档传送给用户浏览。我们采用 JSP 编写了一个解析 VOTable 文档的程序，然后利用解析生成的结果对象生成 HTML 格式文档，传送给用户浏览器显示。实现流程如下图所示



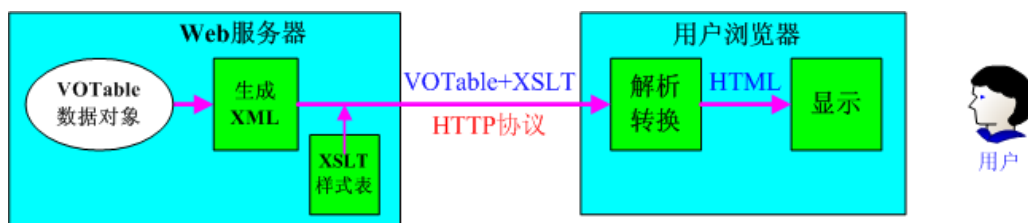
2. 采用我们上节提到的 XSLT 样式表，也就是向 XML 文件中增加一个 XSL 样式表的链接，然后让浏览器来进行转换和显示。这种方式是将网格服务服务调用返回的 VOTable 数据在生成文档时加上一个指向 XSLT 样式表的链接，如指向我们设计的样式表的链接：

```
<?xml-stylesheet type="text/xsl" href="http://www.lamost.org/~sang/VOTable-ChinaVO.xsl"?>
```

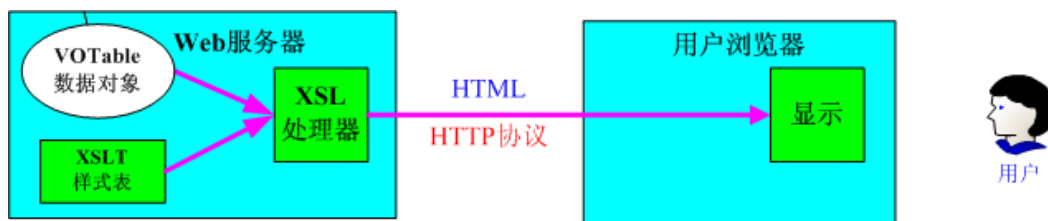
或者指向美国虚拟天文台提供的样式表：

```
<?xml-stylesheet type="text/xsl" href="http://www.us-vo.org/xml/VOTable-basic.xsl"?>
```

然后传送给用户浏览器，用户浏览器根据样式表的规则利用内置的解析器进行实时转换，在浏览器中用户看到的是转换后生成的 HTML 格式文档，但实现前提是浏览器必须支持 XML 和 XSLT 转换。流程如下图所示：



3. 也是采用 XSLT 样式表，在服务器端 XSLT 处理器根据 XSLT 样式表描述的转换规则将 VOTable 文档转换成 HTML 再传给用户浏览器浏览。目前有很多高效的 XSLT 处理器软件包，我们在 Web 服务器端调用 XSLT 处理器，XSLT 处理器根据提供的 XSLT 样式表将 VOTable 文档转换成 HTML，然后传给用户浏览器。流程如下图所示：



如图 5.5 所示为上述第一种 HTML 转换方式显示的界面，图 5.6 为用第二种方式显示的界面。

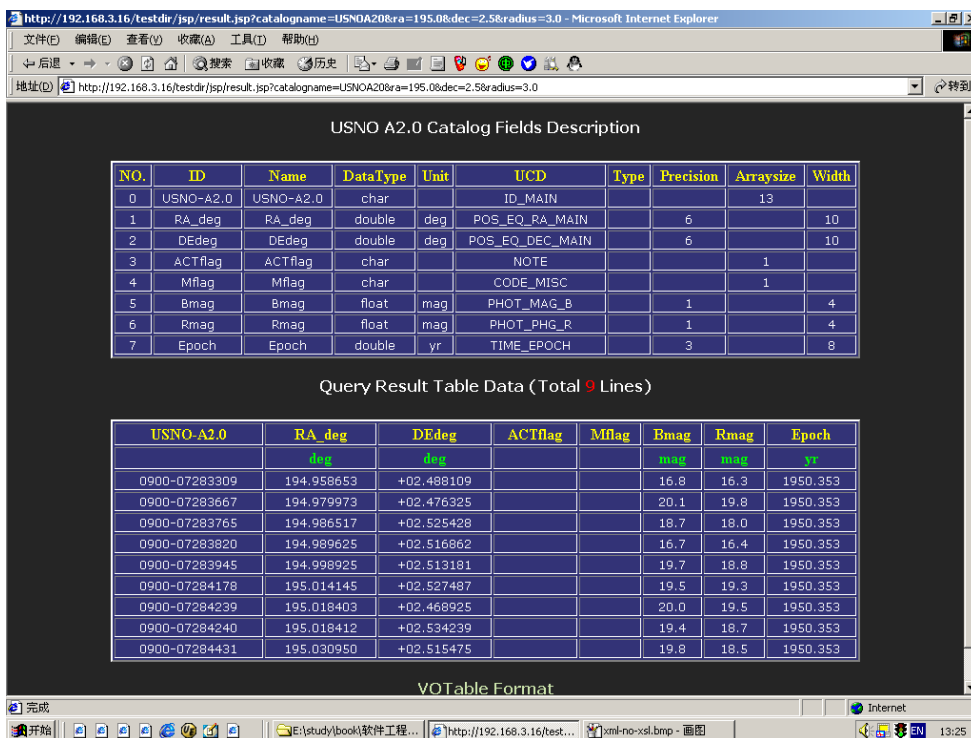


图 5.5: 在服务器端解析 VOTable 生成的 HTML 在浏览器中的显示

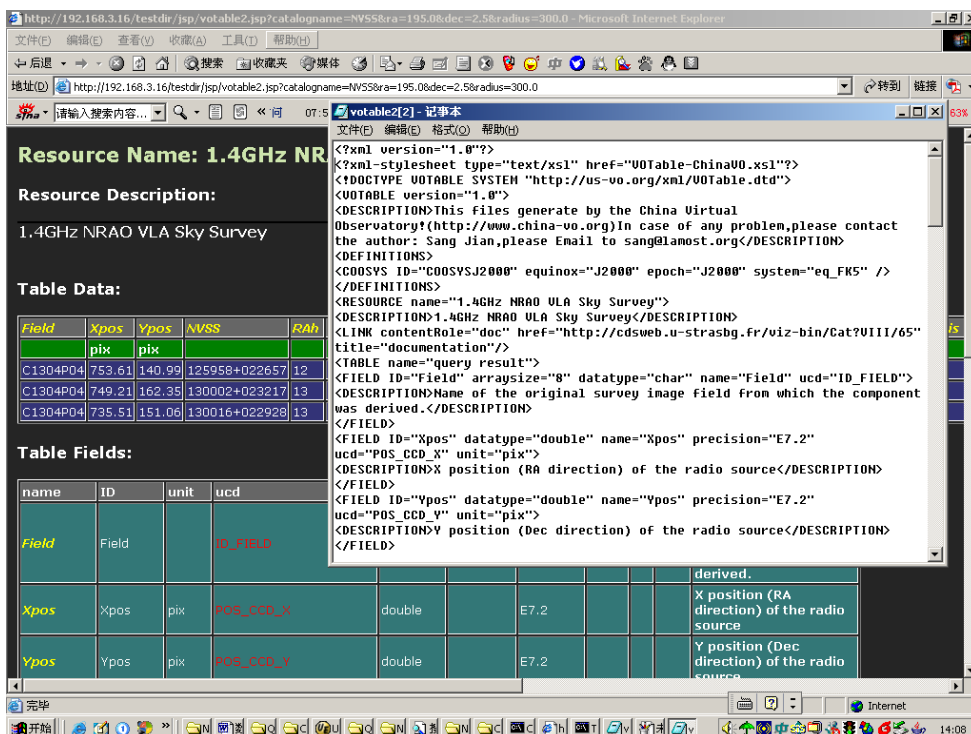


图 5.6: 将包含有 XSLT 样式表的 VOTable 在浏览器中的显示, 右上部分为查看源码的显示结果。



下面对这三种 HTML 显示方式进行比较：

从实现 VOTable 到 HTML 转换的位置来看，第一种和第三种都是在服务器端完成的转换，而第二种是由 Web 浏览器实现的转换。

在 Web 浏览器端实现转换的前提条件是 Web 浏览器必须知道如何执行这种转换，且它必须带有 XSLT 处理器，也就是用户的 Web 浏览器必须支持 XML+XSLT，象 IE 5.5 以上版本，Mozilla，Netscape 6.1 等都内置有 XSLT 处理器（比如 IE 的 MSXML），支持 XSLT 1.0，能实现这种转换。而对于 IE 5.0，虽然支持 XSLT，但其在 XSLT 1.0 最终定案前发布的，仅仅支持 XSLT 工作草案，所以对符合 XSLT 1.0 标准的样式表一定能正确实施转换，而对于更老的 IE、Netscape 则不支持任何转换，但是，对以后新开发的浏览器的基本要求就要支持 XSLT，所以随着新浏览器的推出，采用 Web 浏览器端实现转换方式的通用性会越来越好。采用 Web 浏览器端实现转换的优势是 Web 浏览器将有 XML 和最终的 HTML 文档。如图 5.5，虽然浏览器显示的是 HTML 文档，但查看源代码时还是 VOTable 格式文档，而且实现方式极其简单，如果是引用别人的 XSLT 样式表，所以基本没有什么开发工作量。

在服务端转换的优势是 HTML 在 Web 浏览器中显示的通用性，当 Web 浏览器收到 HTML 时不知道原始文档是 XML，Web 浏览器不需要知道关于 XSLT 或其他如何转换原始 XML 文档内容。这样克服了 Web 浏览器在支持 XML 方面整体上缺乏连续性引起的两难问题，HTML 比 XML 有着更好的向后兼容性，这也就意味着各种用户，包括那些使用低版本 Web 浏览器的用户，都可以通过他们的网络浏览器看到 HTML 文档。其缺点是转换在服务端进行，需要额外的开销进行转换，当服务量大时，会影响到服务的性能。还有一个缺点用户也无从了解用于得到 HTML 文档的 XML 文档结构内容，丢失了 XML 文档结构信息。如果 XSLT 没有提供转换所有信息的规则，那么有可能用户需要的信息恰恰在转换时被忽略掉了。

第一种和第三种的转换实现方式也有其优缺点：

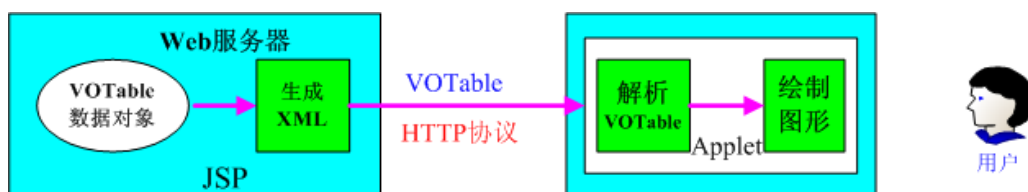
采用第一种编写解析和生成 HTML 的方式来实现转换是一个习惯性的解决方案，XML 的推出要早于 XSLT，所以许多 XML 程序员对 XML 的解析的熟悉程度要远远大于对 XSLT 样式表语法，特别是编写复杂的、功能强大 XSLT 样式表并不是一件容易的事情。但采用 XSLT 样式表转换的方式首先是仅仅需要了解 XSLT 的语法，而不需要接触最低层的 XML 解析的细节实现，减少了编写代码的工作量，

而且也减少了因需要更改 HTML 结果文档结构而带来的代码修改工作量，易于维护和升级。还有一个优势在于 XSLT 处理器的性能得到了很大的提高，采用高性能 XSLT 处理器转换有可能会得到比自己编写的代码更优越的性能速度。更多的优势都源于 XSLT 本身，XSLT 通用性、可重用性好，而且 XSLT 还提供了一个发布 XML 显示规则的机制，像我们编写的转换 VOTable 到 HTML 样式表，只要提供获取地址，那所有的 VOTable 应用都可以使用样式表进行转换。而且 XSLT 不涉及如何实现转换，它仅仅描述 XML 的转换规则，具体转换由 XSLT 处理器完成，具有高度的“实现无关性”，这是编写程序代码不可能实现的（涉及到编程语言，运行环境，接口设计等等）。

### 5.2.3 设计实现以二维图形显示 VOTable 数据的界面

#### 1. 采用 VOPlot 工具包实现 VOTable 的图形可视化

最美观的和直观的可视化 VOTable 的方法是图形显示，而最强大的界面就是交互式界面，我们目前利用印度虚拟天文台（India-VO<sup>[20]</sup>）开发的 VOPlot<sup>[21]</sup>，实现了我们自己的 VOTable 数据图形可视化交互界面。



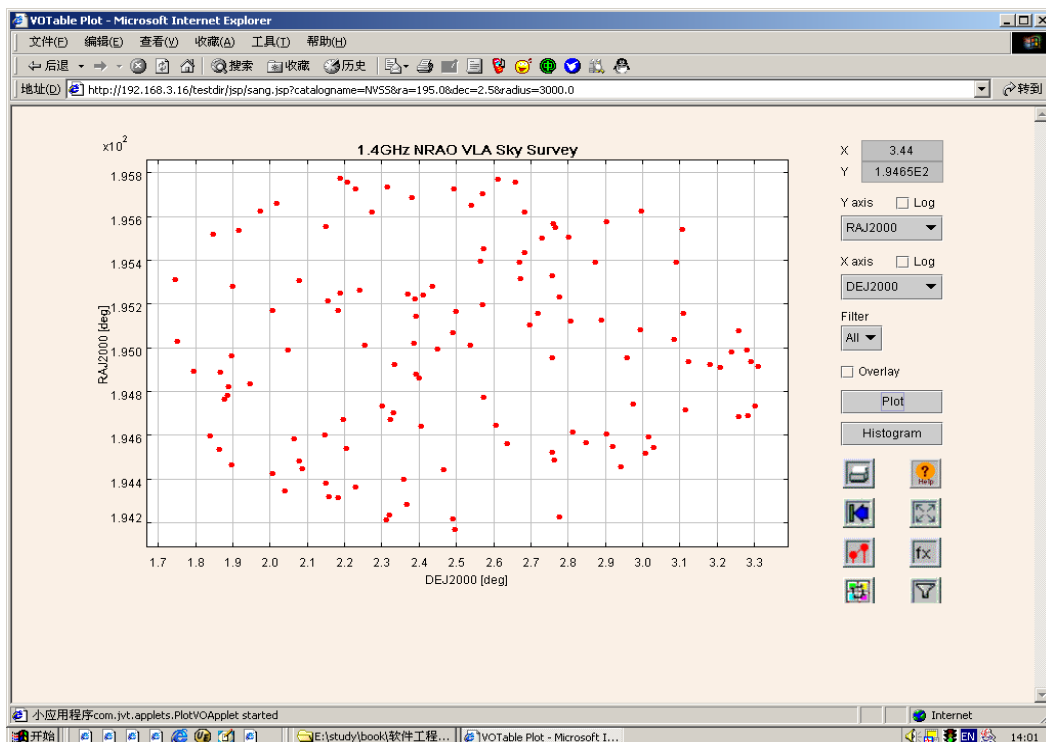


图 5.7: 利用 VOPlot 图形可视化显示 VOTable 数据

VOPlot是VOTable交互式二维可视化工具，它是一个Java Applet<sup>[22]</sup>软件包，能够根据用户的需求对VOTable中的两列数据进行二维显示，VOPlot是目前为止功能最强大的VOTable可视化显示工具之一，采用这个工具大大减少了我们可视化的开发负担。

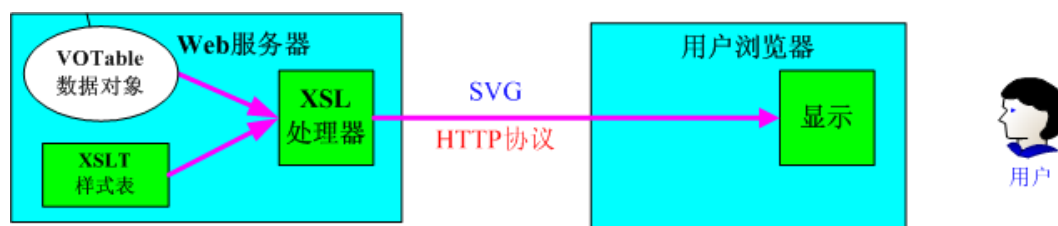
具体显示流程是首先 VOPlot Applet 包被下载到本地 Web 浏览器（浏览器必须支持 Applet）执行后，Applet 去调用 Web 服务器生成 VOTable 文档，VOPlot Applet 在获得 VOTable 后进行解析，然后按照用户选择要显示的列进行二维点的绘制，如图，另外，用户也可以查看 VOTable 数据项的信息。

采用 VOPlot 的 VOTable 可视化界面的优势在于其直观性和交互性，并且容易部署和实现。但其也有缺点，一个是用户的 Web 浏览器必须支持 Applet，VOPlot 大小约为 420KB，在使用时必须临时下载到本地，并且由于解析和绘制都在用户本地进行，需要一定的资源开销。目前版本的 VOPlot 还有一个设计缺陷，就是对包含 PARAM 元素的 VOTable 文档不能正确解析，这造成了一些麻烦。

## 2. SVG 格式的虚拟天图的设想

可缩放矢量图形 (SVG<sup>[23]</sup>) 是一种以 XML 格式描述二维图形的规范, 它包括了创建包含形状和线条信息的 XML 文档的规则和语法, SVG 文档使得用户能够生成一种轻巧的文件格式表示矢量图形, 这些文档的文件名以 .svg 为扩展名结尾, 并且可以由 SVG 浏览器读取显示这些文件, 目前大多数 Web 浏览器支持用插件或 ActiveX 控件查看 SVG 文档<sup>[24,2]</sup>。

像星表的最常见的一种应用就是按照坐标信息和星等信息生成一张虚拟天图, 可以用做观测时的证认图使用。常见的虚拟天图是 JPEG, GIF 等格式, 而 SVG 完全可以描述复杂的虚拟天图, 而且能精确描述天体的相对位置等信息, 所以我们计划采用 XSLT 样式表来将 VOTable 转换成 SVG 格式的虚拟天图。这具有很大的应用前景。但是目前我们没有完成这个 XSLT 样式表的编制, 其可视化实现结构如下: 基本上类似于上述的采用 XSLT 在服务器端转换 HTML 的实现方式, 即根据 XSLT 样式表将 VOTable 转换成 SVG 格式文档, 然后传送给用户浏览器显示绘制。



一种可能的显示如图 5.8:

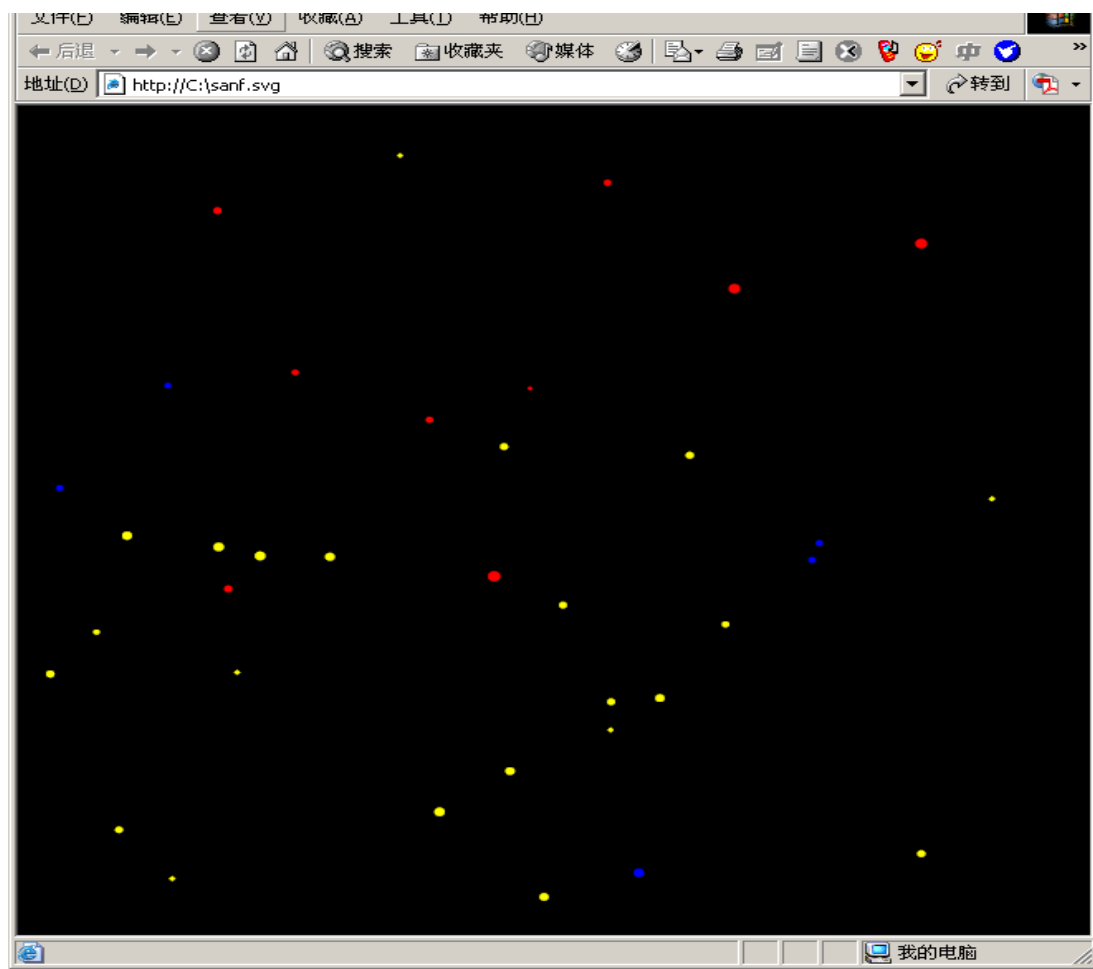


图 5.8: SVG 格式的虚拟天图示例

但是这种方式也有一定的限制，就是 SVG 格式图形与 JPEG，GIF 图像一样，没有 VOPlot Applet 具有那样好的交互性，它是一幅静态图形。但是其优势不可低估，由于 SVG 是矢量图形，与 GIF，JPEG，BMP 格式的光栅图像相比，在缩放旋转时不会降低分辨率和模糊细节，完全可以实现比这些图像更好的显示效果。并且文件轻巧，网络传输开销也要小的多。

## § 5.3 数据可视化服务的设计实现

### 5.3.1 基于网格的数据可视化服务的基本框架结构

我们在上一章提到的数据获取服务都是网格服务，一般的天文学家和公众不可能自己去编写网格客户端，对于他们来说这些网格环境下的数据获取服务是不可见的和不可访问的，所以要为他们提供一个传统的 web 可视化服务查询界面，作为网格和普通网络的接口，使他们真正接入网格。图所示的是天文学家如何通过数据可视化服务等接入网格。

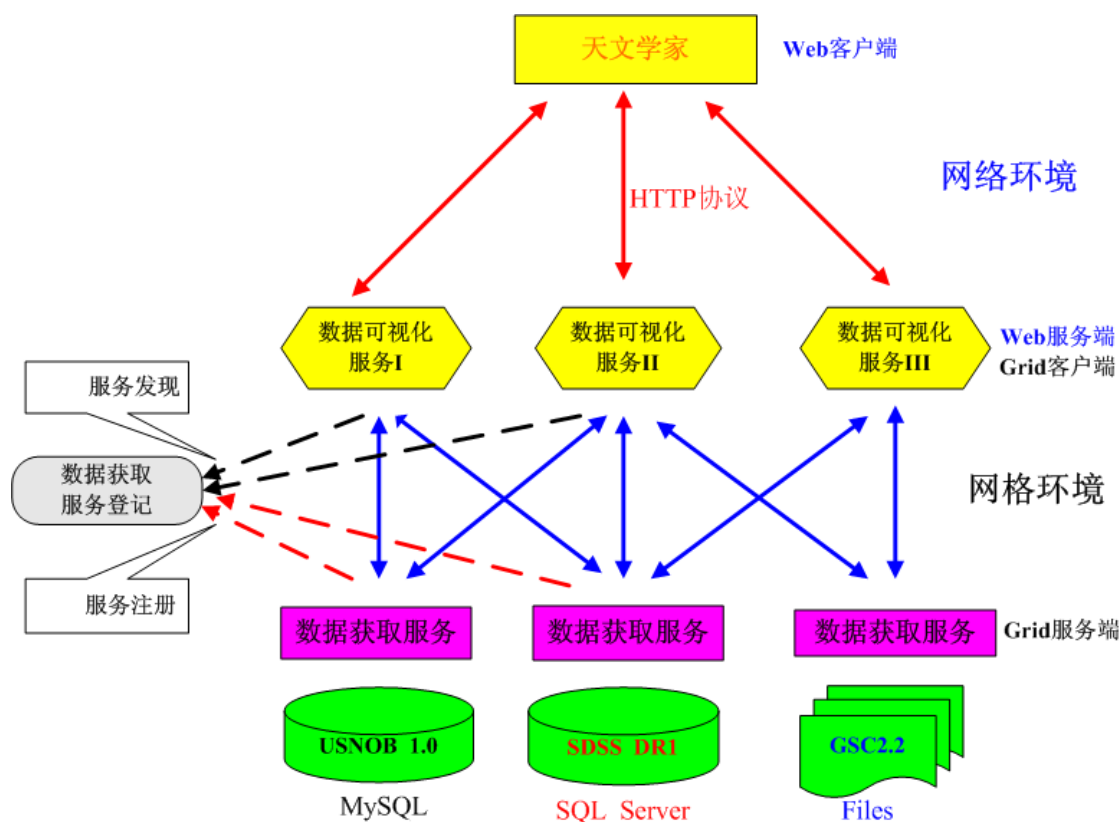


图 5.9: 基于数据服务搭建的数据可视化服务

从图中可以看到，数据可视化服务具有双重身份，对于数据获取服务来说，它是一个网格客户端，调用数据获取服务。而对于天文学家来说，它又是一个 Web 服务器，为天文学家提供 Web 查询可视化界面。通过这个数据可视化查询服务可以访

问到网格环境下的所有数据获取服务。

天文学家通过网络浏览器访问到天文数据的流程如下：天文学家通过网络浏览器发送数据查询请求，一般是通过 HTTP 协议，数据可视化服务接收请求后，去网格服务注册中心查询用户所需要的数据服务的地址，在网格服务注册中心数据获取服务已经注册。数据可视化服务去调用网格数据获取服务，并将返回的数据结果经过可视化处理后返回给天文学家。具体流程在下小节做详细介绍。

我们可以看到以上结构实现了数据服务和可视化服务的分离，不再像传统的数据中心必须为它们的数据集设计一个专门的查询界面。在网格系统中，数据服务不必再考虑数据如何显示，查询界面如何设计，而可视化服务可以专心于显示视觉的设计，而不必考虑对数据库复杂的查询访问。而且更大的益处在于，你可以利用网格内所有的数据服务来为可视化服务提供数据源，甚至这个服务在遥远的什么地方，可视化服务像一个数据 Hub，使天文学家通过一个可视化界面访问到一大批数据服务，这也就实现了数据的统一访问和融合，真正体现了网格的巨大意义。

### 5.3.2 基于网格的数据可视化服务的基本工作流程

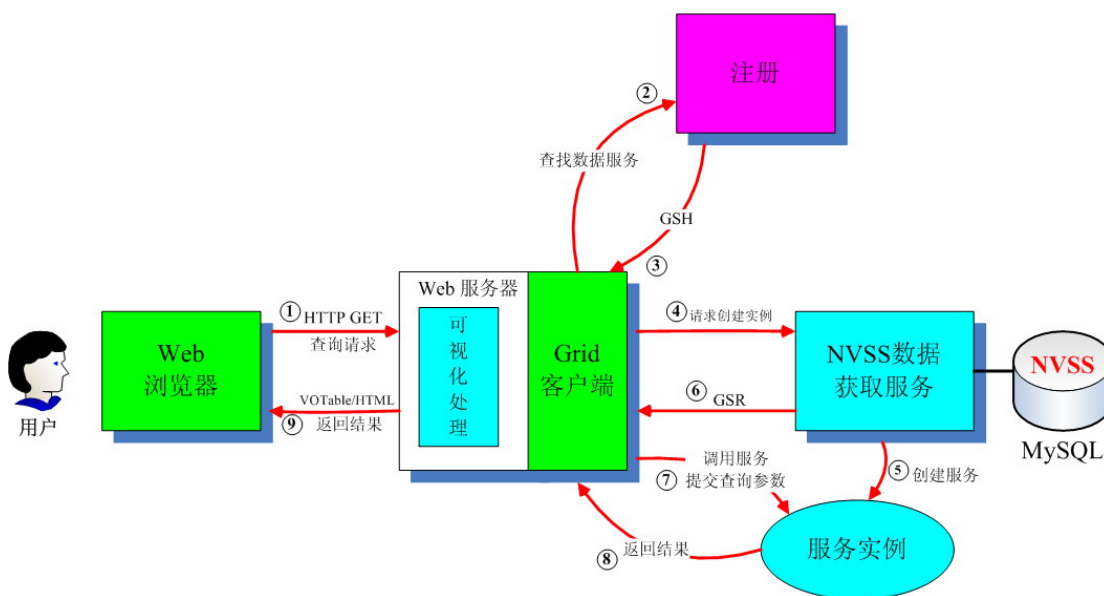


图 5.10: 面向天文学家的数据可视化服务实现流程

基于网格数据服务实现面向天文学家的可视化服务实现的基本流程如图所示，它设计两层的服务实现流程，即面向天文学家的 Web 服务流程和实现自身功能去调用网格数据获取服务的流程。

具体流程如下：

1. 用户通过提交表单发送查询要求，这个查询请求可以通过 HTTP Get 或 Post 方式发送给 Web 服务器；
2. Web 服务器接受到请求后，作为网格客户端去数据服务登记处查询何处能提供用户需要的数据集的获取服务，在此之前，数据获取服务已经向登记处注册服务；
3. 登记处查询到这个服务，并返回服务的 GSH；
4. 网格客户端（Web 服务器）通过登记处提供的 GSH 去访问服务，申请创建服务实例；
5. 数据服务为客户端创建服务实例；
6. 返回创建的实例的 GSR；
7. 网格客户端激活服务实例，通过服务接口调用服务实例去执行服务；
8. 服务实例将查询结果以 VOTable 的格式返回给 Web 服务器；
9. Web 服务器对 VOTable 进行可视化处理，并将结果返回用户浏览器显示。这个可视化处理如前一节描述的实现方式，我们分别采用不同的可视化实现方式创建了不同的可视化服务。

### 5.3.3 我们实现的其中一个数据可视化查询界面的介绍

我们基于上一章提到的实现的一系列星表 Cone-Search 查询服务，实现了一个简单的查询服务界面，在这里可视化实现方式是采用的 VOPlot 图形显示。

Cone-Search 查询界面如下图所示，上边的列表框中列出了可以查询的星表，下面的三个输入框为查询中心点的赤经（RA），赤纬（Dec）和搜索半径（radius），它们分别的单位是度，度和角分。最下面的是提交（Retrieve Catalogue）和重设按钮（Reset Input Fields）。表单的提交方式为 Http GET 方式，即把输入参数附在 URL 后面。



格式如下:

<http://ns2.lamost.org:8080/cvo/result.jsp?catalogname=NVSS&ra=195.0&dec=2.5&radius=30.0>

当用户点击提交按钮后, 用户浏览器向服务器提交的 HTTP GET 消息头

```
GET /cvo/result.jsp?catalogname=string&ra=string&dec=string&radius=string HTTP/1.1
Host: ns2.lamost.org:8080
```

上表的带下划线的 string 代表的是用户输入的真实参数值。

可视化查询服务在得到 HTTP GET 请求后, 去调用网格数据获取服务, 数据获取服务返回给可视化服务的是 VOTable 格式的结果, 然后在用户下载并执行的 VOPlot Applet 解析并显示, 用户可以通过选择要使用的列查看绘制结果。

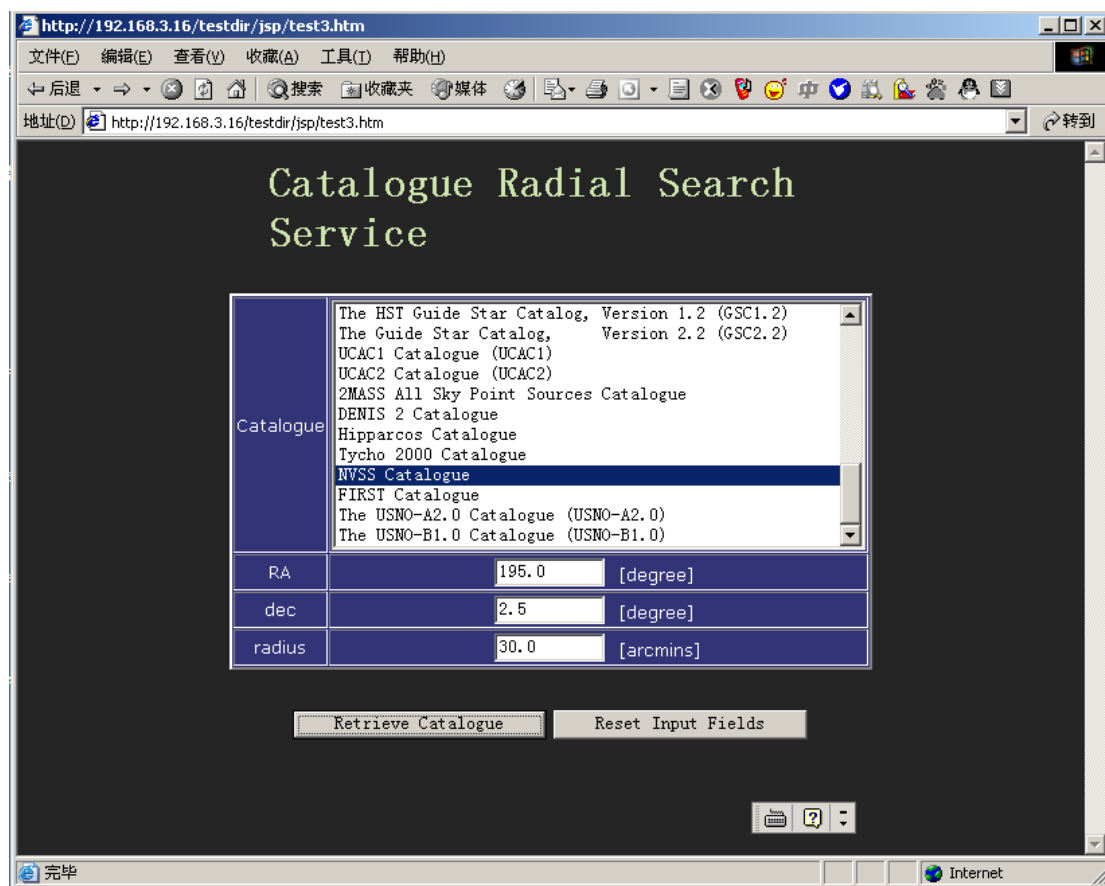


图 5.11: Cone-Search 查询服务的查询界面

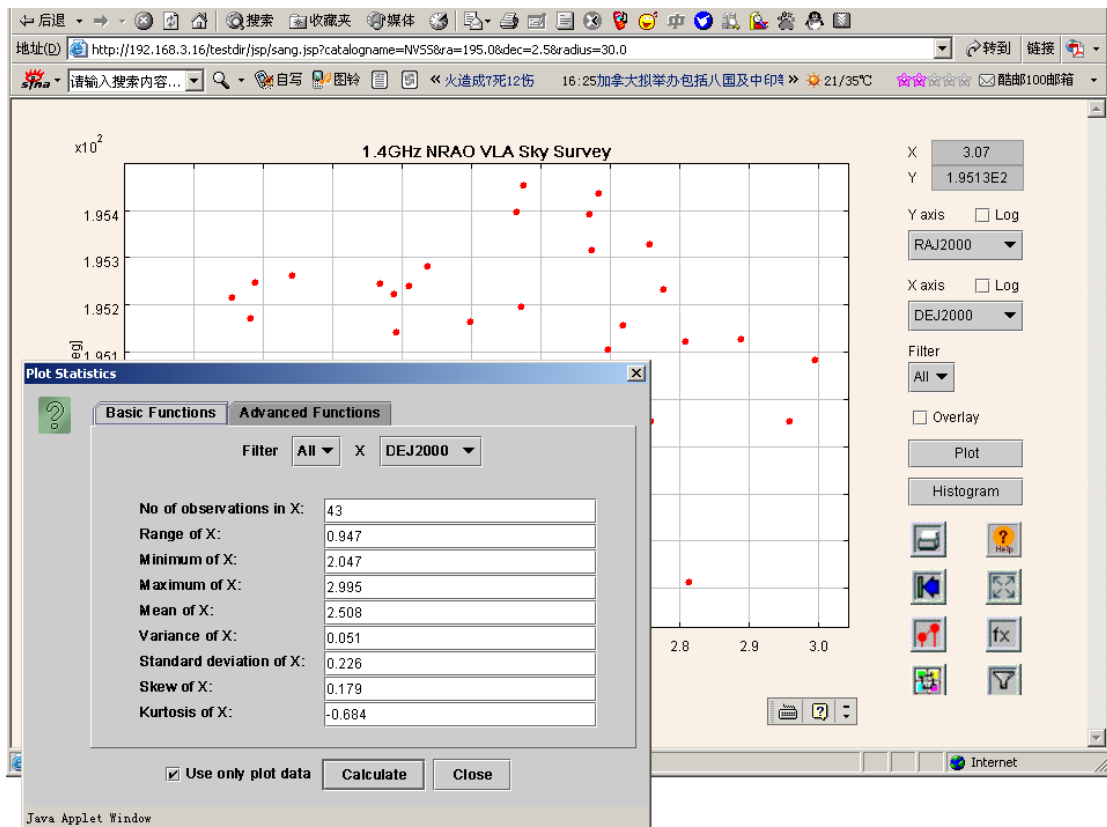


图 5.12: Cone-Search 查询服务的可视化界面

## 参考文献

- [1] [XSLT] XSL Transformations (XSLT) Version 1.0. <http://www.w3.org/TR/xslt>
- [2] Java, XML, and Web Services Bible. Mike Jasnowski. 电子工业出版社, 2002年5月第一版
- [3] Java XML Programmer's Reference, Mohammad Akif, 电子工业出版社. 2002年5月第一版
- [4] XML Bible, 2<sup>nd</sup> Edition, Elliotte RustyHarold, 电子工业出版社. 2002年1月第一版
- [5] XSLT Working with XML and HTML. Khun Yee Fung. 清华大学出版社. 2002年10月第一版
- [6] Inside XSLT. Hoizner Steven. 机械工业出版社. 2002年1月第一版
- [7] XSLT 从入门到精通. 陈建勋. 中国铁道出版社. 2002年1月第一版
- [8] XSLT による文書変換.  
<http://icrouton.as.wakwak.ne.jp/xml/primer/index.xml?section=8>
- [9] Benoit Marchal. XSLT 处理程序是如何工作的.  
<http://www-900.ibm.com/developerWorks/cn/xml/x-xslang/>
- [10] Pramod Kankure. 使用 Java 和 XSLT 生成动态 Web 页面.  
<http://www-900.ibm.com/developerWorks/cn/xml/tips/x-dynweb/index.shtml>
- [11] 乔智君. XML 在空间信息管理中的应用.  
<http://www-900.ibm.com/developerWorks/cn/xml/x-gis/index.shtml>
- [12] Nicholas Chase .技巧: 根据 XML 内容设置 XSL 样式表. <http://www-900.ibm.com/developerWorks/cn/xml/tips/x-tipxsltjx/index.shtml>
- [13] [HTML 4.0] <http://www.w3.org/TR/html4/>
- [14] Paul Grosso, Norman Walsh . XSL Concepts and Practical Use.  
<http://nwalsh.com/docs/tutorials/xsl/xsl/frames.html>
- [15] [SAXON] Open Source SAXON XSLT processor. <http://saxon.sourceforge.net/>
- [16] [Xalan] Xalan-XSLT stylesheet processors, <http://xml.apache.org/>

- [17] XSLT初步. <http://www.zdnet.com.cn/common/printfriendly/printfriendly.htm?AT=39030849-2000081602t-20000562c>
- [18] [MySQL] <http://www.mysql.com>
- [19] [PostgreSQL] <http://www.postgresql.org>
- [20] [VO-India] Virtual Observatory India. <http://vo.iucaa.ernet.in/~voi/>
- [21] [VOPlot] <http://vo.iucaa.ernet.in/~voi/voplot.htm>
- [22] [Applet] <http://java.sun.com/applets/>
- [23] [SVG] Scalable Vector Graphics (SVG). <http://www.w3.org/Graphics/SVG/>
- [24] 网站开发新动力 用 XML 轻松开发 Web 网站. 灯芯工作室. 北京希望电子出版社,2001 年第一版

## 第六章 总结和展望

### §6.1 论文总结

本人从信息时代的天文学的特点和虚拟天文台的产生说起,详细介绍了自己在虚拟天文台研究领域所做的一系列工作。本论文包括的内容有:

1. 第一章介绍了虚拟天文台产生的背景,概念,发展现状和建设中国虚拟天文台的重要意义。
2. 第二章介绍了我们在开发中用到的一些 IVOA 推出的新标准。
3. 第三章介绍了作为中国虚拟天文台的重要基础设施的天文星表数据库的设计和实现,着重介绍了星表数据入库的规则和元数据库的结构设计。并介绍了自动星表数据入库工具的实现。
4. 第四章介绍了我们利用基于开放网格服务架构(OGSA)的 Globus 工具包设计部署的多个数据访问服务。首先介绍星表 Cone-Search 服务接口的设计,以及基于不同的数据库环境实现的多个星表的 Cone-Search 服务。其次,还介绍了实现更强查询功能的星表查询接口的设计,以及基于天文星表数据库实现的星表查询服务。最后本章简单介绍了简单天文图像访问接口的设计以及 DSS-I 天图访问服务的实现。
5. 第五章介绍了自己在数据转换和可视化方面做得一部分工作,其中包括用于将 VOTable 转换成 CSV 和 HTML 的 XSLT 样式表的设计;包括如何在 Web 环境下以 HTML 或图形显示的几种技术解决方式的具体实现,还包括如何基于网格环境下的 Cone-Search 实现面向天文学家的星表查询可视化服务界面。

### §6.2 对未来工作的展望

通过这几年在虚拟天文台方面研发工作的一些体会,我认为中国虚拟天文台的未来发展应该遵循下列的一些原则,以使中国虚拟天文台得到更快更好的发展。

1. 以科学示例带动技术的需求和发展。象中国虚拟天文台这样的小项目，去象 NVO 一样实现一个完整的虚拟天文台网格框架现在是不可能的，应该提出一系列能体现虚拟天文台意义的科学示例，以这些示例作为需求促进技术的发展，一能提高技术发展的速度，二能在发展中得到一定的科学产出。
2. 利用现有的软件资源，开发我们自己的成型工具软件。天文学家是虚拟天文台的最重要用户，我们要为他们提供成熟的软件工具，让他们真正用到虚拟天文台，享受到虚拟天文台带来的便利。在软件工具的开发中，要充分利用现有的软件资源，特别是开放源代码工具软件，这样可以大大减少我们的开发工作量。
3. 保证开发人员工作的衔接性，充分利用前人留下的工作基础。将前人的工作抛弃或者是推翻重来实际上也是巨大的资源浪费。虚拟天文台是跨多学科的研究前沿，需要各个方面的深入，需要形成传帮带的机制，使后来人能很快熟悉研究领域，并能继承前人的工作继续前进。
4. 保持严谨的工作作风，不得有一丝的疏忽敷衍，虚拟天文台也是一项工程，工程的每一个部分都要严格把关，保证任务完成的质量。

中国虚拟天文台现在还处在刚刚起步的阶段，各方面都需要去研究，希望中国虚拟天文台在以下几个方面继续着重进行开发工作：

### ● 天文数据库的建设

天文数据库的建设仍然是未来的工作重点，本人认为应在数据库方面需要继续做的工作：

#### 1. 天文星表数据库的充实和优化；

A) 目前天文星表数据库包含的星表还比较少，仍然需要扩容，需要把天文学家一般用到的星表入库，满足更为广泛的需求，之所以象 VizieR, Simbad, NED 等天文数据库拥有庞大的用户，一个重要的原因就是他们保证了查全率。我们设计的星表入库工具大大减轻了入库的工作量，使得低成本建立庞大的数据库成为可行，希望能继续充实星表数据库。

B) 优化数据库的查询效率。目前数据库仅仅支持 B-Tree 索引，在

基于赤经、赤纬两列的空间查询上性能有限，希望引入一列即可描述坐标位置的 HTM 或 HEALPix 编码方法，可大大提高查询效率。另外需要为数据库建立一系列天文专用的函数和过程，方便使用 SQL 语言描述天文专用查询要求。

2. LAMOST 光谱数据库的设计和实现；我们已经在 LAMOST 光谱数据库的初步逻辑设计上投入较大精力，取得了一定进展，希望继续进行数据库详细物理设计以及整个软硬件系统的实现工作。

3. SkyServer 数据库的迁移工作；SkyServer 数据库对于 LAMOST 具有非常重要的意义，其测光星表可以作为我们进行选源，编制输入星表重要的资源，但 SkyServer 是运行在 Windows 2000 系统的 SQL Server 上，而天文工具和软件一般的运行环境是 Linux 或 Unix，我们的 LAMOST 光谱数据库计划采用的数据库管理系统也是 Unix 下的 PostgreSQL，为方便应用，有必要把 SkyServer 部分表导入到 PostgreSQL 下，这一个是比较容易实现的。

#### ● 虚拟天文台的数据服务的设计与实现

完善虚拟天文台网格服务体系，充实现有的虚拟天文台的数据服务，设计和开发新的天文数据服务也是下一步的工作重点：

1. 完善服务的注册和发现体系；
2. 探讨更为完善的天文图像服务接口的设计，实现更多更强大的天文图像服务。天文图像是天文数据中数据量最大的一部分，如何描述对天文图象的复杂查询要求，设计和实现功能强大的服务接口和功能，也是我们下一步研究的重点；
3. 探讨光谱数据服务的接口设计和实现方式，这是将 LAMOST 光谱数据接入到虚拟天文台的必然需求。

#### ● 数据可视化工具的开发

重视虚拟天文台数据可视化的集成和开发：

1. 我们现在还缺乏一个对 FITS 图像进行各种显示的 Java Applet 工具软件。象 Aladin 这样的图像显示工具不提供源代码，也就无法移植到我们搭建的可视化服务中，我们需要开发这样一个强大的图象显示工具。

### §6.3 结束语

中国虚拟天文台的研究和技术研发还处于刚刚起步阶段,摆在我们前面的道路还很漫长,而且荆棘遍地、困难重重,实现虚拟天文台提出的任务还需要付出更多的努力,但是我相信,通过中国虚拟天文台项目人员的共同努力,通过国内外虚拟天文台同行、天文学家、计算机学家和数理统计学家之间的精诚合作和大力支持,中国虚拟天文台就一定能取得突破性的发展,实现其提出的雄伟目标。



## 致 谢

三年难忘的硕士学习生活即将结束,在此要对许多给过我热心指导帮助的老师同学表示衷心的感谢!

首先感谢我的指导教师——赵永恒,是他指导我——这个既非天文也非计算机专业的大学生走进了天文学和计算机科学交叉的最前沿研究领域——虚拟天文台。虽然赵老师担任着 LAMOST 工程的管理职务,日常工作非常繁忙,但是他仍在百忙之中抽出时间对我进行悉心的指导和帮助,让我在最快的时间了解并热爱上虚拟天文台,并为我指引了明确的研究目标方向。赵老师还组织经常性的小组讨论,为我的研究工作提供了很好的氛围,赵老师还给我们提供了非常优良的计算机设施,这在以前和今后都是不可想象的。赵老师的特点是学术自由,这给予了我很大的自由发展的空间,使我通过三年的学习研究具备了初步的独立科研能力,这对我今后的学习工作奠定了坚实的基础。赵老师渊博的知识、敏锐的思维、远见的卓识、正直的性格和严谨的治学态度不但给我们留下深刻的印象,并在潜移默化的影响着我们。

感谢和我一起进行虚拟天文台项目研究的各位老师同学,他们是:

1. 虚拟天文台项目的统筹者——崔辰州博士,他的知识渊博,思维敏捷,工作组织能力突出,他在我学习工作上给予了大力帮助,并为我们创造了浓厚的学术研究气氛,他认真负责的工作态度值得我们学习。

2. 在天文数据挖掘方向作出突出贡献的张彦霞博士,她待人热心、工作上扎扎实实,她在工作学习上给了我很多有益的指导和帮助。

3. LAMOST 光谱处理的负责人罗阿里博士,他在工作岗位上辛勤耕作,任劳任怨,他是我们共同尊敬的老大哥,他在研究工作中的热心指导使我获益菲浅。

4. 我的工作搭档邵惠娟,正是和她的合作,互相支持下,我们才得以在虚拟天文台方面取得突破进展,她在日常学习生活上给了我很大的帮助,我不会忘记这个好搭档。

5. 我的师妹王小倩,她热心助人,坦诚待人,在工作中我们一起精诚合作,她给了我许多无私慷慨的帮助。正是她的大力协助,使我在最短的时间内完成了毕业论文。

赵瑞珍博士是自动化所的同事中最先认识的，他为人诚恳，待人和善，思维敏捷，他在天文台工作的两年中在生活学习上给了我莫大的照顾，赵大哥，谢谢了！

感谢 LAMOST 项目部和我一起工作学习过的同学们，我们一起工作、讨论问题，一起游玩娱乐，他们给予我无私的帮助和莫大的欢乐，我感觉和他们一块度过的这三年时光是我最快乐、最丰富的一段日子，我很怀念这段和他们在一起的时光。他们是王伟、程林鹏、卞维豪、贾磊、覃冬梅、邱波、王辉；吴潮、朱光华、许馨、刘中田；李博、马忠祥、殷红星、赵梅芳、张健楠、段福庆、李乡儒、杨金福、刘高潮、李绍明。

感谢国家天文台 LAMOST 项目部的老师们，他们是苏洪均老师、石火明老师、陈英老师、袁晖老师、孙盛慈老师、冯磊老师、李有刚老师、门力老师、曹淑蕴老师。

感谢人事处的杜红荣老师，正是杜老师的热情和耐心让我决定来国家天文台攻读硕士。还要感谢杜老师在这三年中的工作生活上给予的大力帮助和关怀。

感谢我们的舍长姜钊和室友马利华、宋文彬、吴潮、徐峻峰、张志勇在生活上的支持与互助。和他们在一起的生活很轻松愉快。

最后感谢我的父母亲和姐姐，亲情是最伟大的，是他们的关爱、照顾和支持使我得以集中精力完成学业。感谢我的父亲，是他给了我“干一行爱一行”的事业理念和严谨的工作作风。感谢我的母亲，是她的自身行动教导我要忘我工作。

## 后 记

作为一个即非天文也非计算机专业毕业的普通大学生，却走进了天文科学和计算机科学的交叉点-虚拟天文台，而且还是走进了交叉科学的最前沿阵地。作为虚拟天文台研发的最早两人之一，在这三年的学习研究历程中，没有前人的技术路线作为指引，其前进之艰难可想而知，在研究之初走了不少弯路，幸于 2003 年 3 月与邵通力合作，做通了虚拟天文台的第一个可行性服务示例——Cone-Search，前景一片光明，然所剩时间已经不多，唯一能做的就是尽自己能力，去一点点的了解，一点点突破，当自己把一些想通，理解透彻，感觉前景一片光明时，正准备为实现自己的思想大展宏图时，时限已到，毕业——这一道截止线却逼迫自己放弃实现这些自己的想法，自己新的研发目标，自己许多的想法和工作没有最终完成。总结自己三年来的工作学习，认为自己虽然愚笨，基础差，但自己尽力去弥补在天文学、网格技术等各方面的不足，我认为自己没有浪费这三年的宝贵时光，并取得了一定的成绩，具备了初步独立研究的能力。

在对待虚拟天文台研究工作上，个人已经把虚拟天文台的研发作为一项兴趣去做，兴趣—这是个人执著之所在，自己得以全身心投入其中，至现在仍难以“自拔”，实际上自己取得的一些进展源于自己兴趣引发的好奇心，并在进行一些小尝试后发现了极大的应用前景。希望每一位从事中国虚拟天文台研发的同学，都能把兴趣和热情投入到其中来，为中国虚拟天文台的发展贡献自己的力量。