

doi: 10.3969/j.issn.1000-8349.2013.01.01

天文数据库回顾与展望

李 建, 崔辰州, 何勃亮, 赵永恒, 曹子皇,
樊东卫, 李长华, 谌 悦

(中国科学院 国家天文台 光学天文重点实验室, 北京 100012)

摘要: 随着新一代巡天观测、时域观测等天文项目的推进, 当前的天文数据量越来越大。面对天文领域日益增长的大数据集和大数据流, 需要一些高效的数据存储和数据处理方法来对数据进行管理。天文数据库以及在其基础上提供的数据服务的出现部分解决了这样的问题。借助于计算机技术和信息技术的进步, 一些更加专业的针对天文领域的发现、数据挖掘、数据交互等工作正在逐步标准化。基于虚拟天文台技术和天文信息学的新型天文数据服务正深入天文用户的日常科研生活中。首先简单回顾了天文数据库的历史, 然后通过对几个典型天文数据库的举例分析, 从天文数据库的类型、提供的服务等方面介绍了当前天文数据库的特点和最新进展, 并对今后天文数据库的发展做了展望。

关键词: 天文数据库; 数据检索; 交叉认证; 巡天观测; 虚拟天文台

中图分类号: P1, TP392 **文献标识码:** A

1 引 言

从古至今, 人类对宇宙的探索已经留下了形式多样的天文数据。当代天文数据主要有: 记录天体位置和天体参数等信息的“星表数据”, 直接对天空某一区域进行拍摄的“图像数据”, 反映目标天体重要物理信息的“光谱数据”, 时域天文观测产生的“时序数据”, 进行计算机模拟天体演化等现象产生的“数值模拟数据”以及天文学家留下的科学文章、文献等。

计算机技术、CCD 相机技术应用于天文领域后, 天文数据开始实现数字化存储。为了统一数据格式, 天文学界制定了天文数据的标准。FITS (Flexible Image Transport System)^[1] 是国际天文学联合会 (IAU)1982 年确定的世界各天文台之间用于数据传输、交换的统一标准格式。目前大部分主流的天文数据都为 FITS 格式。FITS 文件由文件头和数据组成, 在文件头中有对该文件的描述, 如观测时间、观测对象、曝光时间等信息, 非常方便科研人员对其进行查看和提取。

收稿日期: 2012-7-12; 修回日期: 2012-12-19

资助项目: 国家自然科学基金 (10820002, 60920010, 90912005); 科技部国家科技基础条件平台建设项目 (BSDN2009-07); 北京市科技新星计划 (2007A085)

国际上,近年来大规模图像巡天和大样本光谱巡天方面已经取得了长足的进展^[2],特别是一系列光谱巡天计划的成功实施,使人们获得了空前丰富的星系和恒星光谱资料,推动了天文学各个分支的蓬勃发展。同时,它们产出的天文数据也在逐年递增。另外,国外一些国家未来几年即将建成的几个大的天文项目如 SKA^[3]、LSST^[4]、Pan-STARRS^[5]、ALMA^[6]、SDO^[7]、ATST^[8]等提出的存储数据、处理数据的要求也都十分巨大。

面对庞大的天文数据量,对天文数据进行系统的电子化归档,无疑会极大提高工作效率,尤其是利用数据库技术进行管理后,使得星表搜索、认证等工作有了全新的操作方式。成熟的数据库技术出现在天文数据大规模电子化管理之后,然而早期的天文学家已经意识到有计划地利用电子计算机管理天文数据对未来天文研究的重要性,因此开始了天文数据库早期的探索性工作。数据库技术的出现,使天文学家有了更为便捷的管理工具,他们开始在数据库的基础上,实现各种天文数据的科学应用:为了满足天文数据特殊的检索需求,开发出了高效的锥形检索方法^[9];为了实现数据集之间的匹配问题,开发了交叉认证工具^[10-12];为了更直观地提供天文数据服务,开发了各种可视化工具^[13, 14];……。目前已经形成了数目可观的天文数据库服务,使整个天文数据管理服务形成了一个良好的“生态系统”。本文在第二章回顾天文数据库的历史,展现数据库的发展历程;然后在接下来的第三、四章中,从天文数据库分类和天文数据库所能提供的服务两个方面,详细地介绍当前天文数据库的特点;最后,针对处于当今日新月异的计算机技术和信息技术环境下,天文数据库的应用和未来发展做了展望。

2 天文数据库的历史

当前,天文学家尤其是刚刚进入天文学领域的研究生已经对在互联网上检索、下载、处理天文数据的工作方式习以为常。然而天文数据从传统的纸张记录到现在的完全电子化存档管理,经历了一个漫长的过程。纵观天文学发展的历史,越到近代,相同时间段所产生的科学数据的量就越大,也就越需要系统化的存档管理。电子计算机的出现不仅为处理数据提供了崭新的途径,而且为天文数据管理带来了新的方法。不仅如此,数据库在天文上的应用,也促进了数据库技术本身的发展,面对未来海量的天文数据,数据库领域不得不重新考虑开发一种更高效、合理的新型数据库,以解决搜索速度、数据处理等所面临的技术瓶颈问题。

2.1 第一阶段——天文数据库起步

20世纪70年代以前,计算机技术正处于起步阶段,少数的一些大型计算机也只是应用于军事、科学计算方面。数据多数采取纸带、打孔卡的存储方式。计算机存储并没有显示出明显的优势,基于计算机的数据库概念正处于酝酿阶段。20世纪七八十年代,微型计算机开始在商业上应用,早期出现的数据库主要用在电子财务和商品管理等数据量、访问量不是特别大的应用上。此时,一些数据管理方面的天文学家开始意识到利用计算机存储、管理天文数据的趋势。但是此时并没有特别适合于数据量庞大的天文数据的数据库,于是一些天文学

家开始编写适合自己天文台的数据存储程序, 然而这还称不上是一个数据库系统, 只是稍复杂的基于文件系统的数据存储管理。

法国斯特拉斯堡天文台的 CDS 数据库是天文数据库领域的先驱^[15]。它在 1972 年, 就开始采用计算机进行电子化管理天文数据和文档。最初只是利用电子表单搜集“有用的”天文数据, 并且加以整理、归档。早期只限于内部使用, 或者国际合作伙伴的使用。随着互联网技术的发展, CDS 把数据开放到互联网上, 才开始为更广泛的用户提供天文数据和服务。

CDS 从 20 世纪 70 年代建立开始就提供 Simbad^[16] 服务。Simbad 是一个太阳系外的天体目标数据库, 到目前为止一直不间断地更新。它能够为用户提供太阳系外所有天体目标的一些基本参数、别名和相关文献, 是一个非常不错的天体目标交叉证认工具。在天文学中, 同一目标的天体经常有不同的叫法或命名规则, 不同天文台在整理观测结果时设计出的星表也有自己的编号体系, 因此一直以来天体目标的交叉证认工作就是一项意义重大但又耗时耗力的重复性工作。这样的工作对计算机来说, 最适合不过: 对于有规律可循的大量数据, 很方便使用一套程序来管理。Simbad 的成功让天文学家看到了计算机管理天文数据的优势。

2.2 第二阶段——基于关系型数据库的天文数据管理的普及

每一次观测技术方法的改进, 都会使天文学向前迈进一大步。20 世纪 80 年代后期, 随着各种电子元器件尤其是 CCD 技术的成熟和广泛应用, 产生了现代的天文观测手段, 望远镜所得到的观测数据直接就是电子文档。这也为计算机管理天文数据提供了便利。

20 世纪 80 年代初, 关系型数据库初步被人们接受。到 20 世纪 90 年代以后, 关系型数据库的概念已经成熟, 业界也出现了许多优秀的关系型数据库产品如 Oracle、SQL Server、Sybase、DB2 以及后来的 Mysql、PostgreSQL 等。这一时期, 互联网规模和网络传输速度迅速增长, 到 2000 年以后, 已经方方面面影响着人们的日常生活。天文数据库也不例外, 借助新的数据库技术、互联网平台和电子邮件 (E-mail)、网络门户 (Web Portal)、文件传输协议 (FTP) 等各种网络发布手段, 使得原来只能在本地应用的数据库服务普及到网络上, 再加上一些专门开发的可视化网页应用程序, 用户可以以非常友好的方式在世界各地访问需要的数据源。

时至今日, 已经产生了各种基于关系型数据库的天文数据库平台。NED、MAST、ADS、CDS、SDSS、2MASS 等一大批影响力很大的天文数据库正 24 小时不间断地向天文学家开放。尤其在 20 世纪 90 年代后期, 出现了各种大型的巡天项目, 这些项目的共同特点就是会产生大量的观测数据, 基于关系型数据库的天文数据管理无疑在这方面具有巨大的优势。

此时 CDS 的 Simbad 服务已经完全使用数据库管理, 这使得查询、对比、插入、修改数据信息更加方便灵活。CDS 也在这一时期推出另外一项服务——VizieR^[17]。VizieR 是放眼于世界范围内的一个星表集合, 并且融合了和星表相关的图像数据、光谱数据等。它把许多天文星表组织起来成为一个大的数据库, 以统一的标准对外提供服务。到目前为止, VizieR 整合的星表数据集已经超过 1000 个, 星表超过 10000 个。Simbad 和 VizieR 两个服务互为补充: Simbad 重在搜集天文文献中的天体目标, 并且有组织地管理; VizieR 则重在科学数据的管理和维护, 包含了许多完备的大型巡天星表和数据。可以说, 没有先进的计算机数据

库技术, CDS 无法做出如此出色的工作。

另一个典型的例子就是斯隆数字巡天数据库 (SDSS)^[18]。SDSS 项目组和微软的合作伙伴利用 SQL Server 数据库, 针对 SDSS 的数据开发出了一套深受天文学家好评的天文数据服务平台——Skyserver^[19]。Skyserver 以其快速的查询速度, 方便的批量下载数据服务, 灵活的 SQL 搜索以及直观的可视化界面深受用户喜爱。目前为止, SDSS 项目已经稳定运行 10 余年, 释放出 8 批次的数据库。SDSS 的数据已经在各种天文出版物中被广泛引用, 成为目前国际天文界研究最多的数据集之一。而这一切和 SDSS 背后的数据库服务支撑平台密不可分。据统计, SDSS 在软件方面的投入超过总投入的三分之一。后来许多天文项目的数据库发布模式也都深受 SDSS 的影响。

对天文学领域贡献极大的机构是美国的国家航空与航天局 (NASA)。NASA 不仅有一套专业管理天文数据的系统, 而且贡献了很多天文领域的文件标准和数据处理软件。NASA 的数据覆盖面比较全, 涉及到各个波段、各个天文研究领域。因此 NASA 的数据管理方式就是按照各个研究领域或是天文项目 (尤其是太空项目) 来组织的, 例如高能天体物理科学数据库研究中心 (HEASARC)、STScI 多任务数据库 (MAST)、NASA 行星数据系统 (PDS) 等。另外 NASA 的一些数据库也以挂靠到合作单位的方式共同建设和维护, 如美国加州理工 NASA/IPAC 河外数据库 (NED)、NASA/IPAC 红外亚毫米数据中心 (IRSA)、美国航空与航天局/哈佛史密松天体物理数据中心 (ADS) 等。所有 NASA 下属的数据服务, 构成了一套百科全书式的体系, 专业而又齐全。

近年, 尤其是 2000 年以后, 天文数据库蓬勃发展, 除了上面所述的天文数据库外, 欧洲南方天文台 (ESO) 数据库^[20]、加拿大的天文数据中心 (CADM)^[21] 等天文数据库都各具特色。这些数据库除了提供数据搜索、数据获取的服务外, 还都尽量提供和天文学家相关的简单数据处理服务, 使天文数据库网站形成了一个完整的数据服务平台。随着计算机技术的不断进步, 各家天文数据库也都纷纷应用最新的技术来完善自己的数据服务平台, 包括早期的天文数据服务元老 CDS 也开发出了一套出色的可视化平台 Aladin^[14]。

2.3 第三阶段——超大规模天文数据集对数据库的挑战

随着微电子器件工艺日趋成熟, 成本更低、分辨率更高、读出速度更快的 CCD 相机已经广泛应用于各种天文观测中。硬件技术的进步, 使天文学家开始规划天区范围更广、深度更深的巡天项目。另外, 一些较新的天文研究领域如伽马暴、超新星爆发等, 使得时域天文观测的需求更加迫切。所有这些科学需求, 将直接导致天文数据量呈爆炸式的增长^[22]。美国天文最新的 10 年发展规划 (Astro2010) 推荐的第一优先项目——大口径全景时序巡天望远镜 (LSST), 就是在这种背景下提出来的。它将以极高的巡天效率完成更广、更快、更深的巡天观测, 对宇宙学、星系科学、银河系结构、时域天文等诸多方向的研究将起到重要的推动作用^[23]。按照设计要求, LSST 要在每 15 s 得到一个 6 GB 的天区图, 在 5 s 以内快速处理完生成的数据, 按照每晚产生 2000 幅图像来算, 每晚产生的数据量就达到 30 TB。这些数据每天还要从南美洲传回美国本土, 如此工作方式要持续约 10 年。所以 LSST 建成后, 会形成一个图像总量达 100 PB~200 PB、星表总量达 20 PB~40 PB 量级的超大规模的数据库。LSST 在存储能力、数据处理能力、IO 带宽通信能力方面的要求, 无论是在当前还是未

来几年都面临技术上的挑战。除此之外, Pan-STARRS、SKA、ALMA 等正在进行中的新一代天文项目^[24]也具有相似的需求。

为了适应天文数据使用的要求, 一些现代数据库技术, 如索引查询、触发机制、事务处理等功能被用到了极致。例如 SDSS 为了提高搜索速度, 数据库内大量建立索引、视图, 其所占的存储量已经达到数据本身的量级; 并且深度利用缓存机制, 采取各种优化方式, 最大程度地利用整个数据服务器的硬件性能。虽然如此, 当前用在天文方面的主流关系型数据库, 如 Oracle、Mysql、PostgreSQL、SQL Server, 在处理像 LSST 这样量级的数据时依然显得力不从心。天文领域的“大数据”时代, 即将随着 LSST、Pan-STARRS、SKA、ALMA 等高数据产出的项目而到来。届时, 基于海量天文数据的大数据, 在大规模分布式并行数据库、云计算平台等新一代计算机技术支持下, 将实现更加强大的数据访问和数据共享能力。

天文数据最初是借用先进的计算机技术来进行管理的, 然而天文方面强大的需求也促进了数据库技术领域新的思考^[25]。当传统的关系型数据库已经不再适应未来的需求时, 一种全新概念的数据库 SciDB^[26] 在 LSST 需求的启发下诞生了。它突破了现在的关系型数据库模型, 使得 SciDB 的设计更符合科学数据管理的要求。另外, 先进的设计理念可以把数据信息和数据体同时存入数据库, 以方便后期的数据处理过程。可以想象, 如果未来的许多天文数据处理都能在数据库中完成, 那么数据的交叉认证、数据融合等工作将会十分方便。未来天文数据库的趋势是把更多的科学相关的处理服务都融入进来, SciDB 的存储方式无疑具有很大的优势。

3 天文数据库的类型

天文数据和其它学科的数据相比有显著的特点。天文数据不是传统意义上的文本文件、图像文件, 它有自己的专用格式; 另外天文数据向全世界范围开放, 任何人都可以免费获得。随着各国天文项目的不断增多, 天文数据总量也越来越大。天文数据按照观测任务可以分为巡天观测数据和定点观测数据等; 按照观测波段可以分为射电、红外、光学、紫外、X、 γ 等波段数据; 按照数据性质可以分为图像数据、光谱数据、星表数据、时序数据、数值模拟数据等。然而如何组织和规划这些天文数据, 每个数据库站点都有自己的想法, 不尽相同。从不同的天文数据库发布的网站可以看出天文数据库有以下一些类型。

3.1 按观测任务、项目发布的数据库

由于每一台望远镜或是每一个天文项目, 自建立起都有自己的科学目标和观测计划, 所以由项目组按照望远镜观测任务的形式发布一个数据库, 是理所当然的想法, 这也是当今大多数天文项目发布数据的方式。这种形式的数据库优点是显而易见的: 数据集格式统一, 发布人员十分熟悉数据内容, 易于日常的数据维护。一个缺陷就是和其它项目数据集联系不够紧密, 使用者如果需要同一目标在其它观测任务中的信息时, 需要自己去做交叉认证的工作, 增加了使用难度。因此, 有的项目数据库在设计时, 就考虑到和其它相关数据库进行关联, 最大程度地方便用户。斯隆巡天的 SDSS 数据库和两微米巡天的 2MASS 数据库都是这种基于

项目而建立起来的数据库。

3.2 按研究领域、研究机构发布的数据库

天文数据另外一种组织方式,是按照研究领域而建的数据库。类似于 NASA 这样庞大的机构,它所涉及的领域是全方位的,总共的项目和任务成百上千,所有这些任务都是针对几个、十几个研究领域而制定的。因此把所有相同领域的数据库都汇总起来,统一提供几个领域的数据库是其最佳的管理方式。例如 NASA 的 HEASARC 覆盖了 WMAP、COBE、IRAS 等任务;IRSA 汇集了 BLAST、COSMOS、2MASS、Spitzer、Planck、WISE 等观测任务;MAST 包含了 HST、FUSE、GALEX、XMM 等多个任务不同波段的图像和光谱数据。类似这种方式组织天文数据的还有 ESO 的科学数据库、CADM、日本国立天文台的数据中心 (NAOJ) 等。可以看出这种按照领域而分类的数据管理方式,都是在天文实力较强的研究机构,它们有能力开展多项观测任务。这种数据管理方式使其更方便地对数据统一管理和控制。

3.3 文章、文档型数据库

天文数据除了观测所得到的直接用于科研工作的数据外,另一个很重要的资料就是科学文献数据,科学文献是前人工作的结晶。天文学家在开展一项新的科学研究时总是要参考前人的工作,在其基础上进行新的科学探索。ADS 是天文学家广泛使用的文献查询系统。它由 NASA 与其合作伙伴哈佛史密松天体物理中心共同开发和维护。它主要包含了 Astronomy & Astrophysics、Physics & Geophysics 和 arXiv.org 预印本的天文部分内容。用户可以方便地通过作者、天体名称、日期范围、标题关键词、摘要关键词等在 ADS 查找自己所需要的文献。ADS 是专门为天文研究人员所开发的数据库,除了和其它学科文献搜索系统通用的一般功能外,ADS 还具有一些专门针对天文用户的使用模块:如搜索结果集成了 SIMBAD、NED 的服务,可以通过直接搜索结果中的链接,找出文章中所涉及到的所有天体的交叉认证信息;再比如 ADS 还集成了文章阅读的统计信息,使用户了解其他人对本文的关注程度等。

3.4 综合服务型数据库

还有一类天文数据库,它们自己并不产出数据,或者是并不仅仅关注于自己产出的数据,而是把主要的精力投入到所有天文数据的服务上。例如具有悠久历史的法国斯特拉斯堡天文台的 CDS,到目前为止已经可以提供 Simbad、VizieR、Aladin 三大服务。Simbad 数据库用于所有天体目标的交叉认证;VizieR 是包含了几乎所有星表集合的大数据库;而 Aladin 是天文数据的可视化查询、分析工具。除此之外,还有一些非常灵活的在线使用小工具,它们互为补充,涵盖了天文数据处理的主要应用。所以 CDS 的这种服务型的天文数据库,为其它天文数据库站点提供很多启示。

4 天文数据库的服务

前面已经提到了各种类型的天文数据,而对每一种类型的数据,天文数据库都有其对应的服务,如星表服务、图像服务、光谱服务等;并且,每种服务都有其对应的查询、预览、下载以及简单数据处理等方案。除了传统的光学波段的数据外,现代的天文数据来源还扩展到

射电、红外、紫外、X、 γ 等波段; 因此, 同一目标在不同波段的数据融合工作也是天文学家目前比较关注的研究内容。多波段数据的无缝融合工作在现代天文研究中显得很重要。天文技术人员始终致力于如何把网络数据与本地数据的无缝连接、本地软件与网络服务的连接、软件间的协同工作、资源与服务的发现等方面取得突破, 以便在尽量不改变天文学家现有研究方式和工作习惯的前提下, 把丰富的数据资源和先进的网络服务带给他们。

在前面的章节中, 我们已经简单介绍了一些当前比较成熟的天文数据库。从这些典型的天文数据库中, 我们可以归纳出一些现代天文数据库的特点。下面我们就从数据库运行与发布平台、数据库的检索服务、数据库的下载服务、数据库的交叉认证以及多波段融合等方面, 系统地介绍现代主流天文数据库的特点。

4.1 数据库运行与发布平台

和所有其它领域内的数据库相同, 天文数据库可以架设在 Windows 平台或 Linux 等类 Unix 平台上。近些年, Linux 系统服务器的稳定性已经得到公认, 再加上 Linux 系统本身的开源特性, 开发人员只需要很小的成本便可以架设性能稳定的服务器; 所以大多数科研、教育领域的数据库服务器、网页服务器、邮件服务器等都部署在 Linux 服务器上。同样, 许多数据库产品也都可以部署在 Linux 平台上: 用于大型专业数据管理的 Oracle, 小巧轻便的 Mysql, 性能严谨的 PostgreSQL 等基本可以满足天文数据的需求。目前很多天文数据库系统是用 PostgreSQL 来管理的。Windows 系统下的天文数据库并不是没有, 具有极佳口碑的 SDSS 的 Skyserver 系统就是在 Windows 平台下的 SQL Server 部署的, 无论在查询性能、下载方式还是界面系统方面都是天文数据库的典范。

天文数据库一般是利用万维网站服务向用户提供数据和服务的, 因此, 天文数据库一般都需要一个前端网页界面。天文数据库的发布网站不同于普通的商业网站, 它重在简单、实用, 不需要太多的花哨吸引眼球的元素, 清新、大方的风格更能让用户接受。然而网站布局是否安排得清晰、明朗, 会影响到用户的使用。天文数据库网站面向的主要使用群体为天文学家、天文专业的研究生等。由于天文数据自身的特点 (即涉及到不同波段的星表, 图像, 光谱的查询、下载、简单处理等问题), 如何在网站页面上尽可能地合理排布、分列这些功能是一个很有技巧性的工作。前面所列出的天文数据库网站的例子, 如 CDS、NED、SDSS 等都有自己的一套设计逻辑, 用户在体验网站功能时能够快速地上手操作。

除了专业用户的需求, 也要适当考虑到一些大众和天文爱好者用户的使用要求。发布数据库的网站可以使用一些更直观的图标按钮, 进行可视化的操作来减小普通用户的使用难度。网站中应力求更少的专业术语, 如果需要也可以在适当的地方提供必要的解释或帮助链接。实际的经验表明, 完整的使用说明或是例子, 不仅对天文爱好者不可或缺, 即使是对专业的天文科研人员也是十分必要的。

4.2 数据库的检索服务

天文数据中图像、光谱数据本身的数据体都比较大, 因此通常情况下并不是把数据本身存入数据库中, 而是把每一幅图像、每一条光谱的元信息提取出来, 连同存储这些数据的路径存入数据库中。这样做, 有利于高效地使用数据库强大的增、删、改、查等功能对数据进行

操作。由于天文数据一般都是用 FITS 文件存储,而 FITS 文件的文件头就是用来记录如观测目标、观测时刻、曝光时间、天气状况、目标参数等信息,所以只需要把对应的 FITS 文件头信息提取出来,再加上存储路径等信息录入到数据库中,就可以方便地实现信息入库。

针对天文数据量大、类型多等特点,其查找形式也多种多样。最直接的搜索就是针对某一特定的天体目标进行查询:用户输入天体目标名称,或是赤经赤纬坐标,提交到数据库后台处理,然后网站返回所查天体目标的信息。然而,根据天文数据的特点,用户往往不只关心某一特定目标的天体,还需要以某一天体为中心,查找周围一定范围内的所有天体。针对这种情况,技术人员开发了“锥形检索”服务。锥形检索需要用户输入天体目标的位置和搜索半径(通常是以角秒为单位),经过后台的一系列处理,就会输出所选定天区所有目标的信息。

除了这种按照“用户需求”思路的搜索方式外,按照数据本身或数据库本身的逻辑进行浏览,是另外一种重要的数据检索方式。例如每个数据集的数据都是望远镜按照时间顺序观测得到的,建立一个按照数据的观测时间为顺序的浏览数据的方式,就更能使用户清楚数据在时间上的关联;数据库在组织数据时都是按照一定的规律进行的,具有一个唯一的 ID 标识,往往在批量处理数据时,这种按照 ID 查询的方式更有优势。除此之外,针对特定的观测项目还会有一些个性化的检索方式,如按照观测流水号、仪器号、观测天区号、目标星等、数据质量(信噪比、望远镜的运行状况、天气状况、数据处理状况等)的搜索,以及更加直接的数据库的 SQL 语句接口检索。所有这些方式的搜索,其核心目的都是使用户最方便地查询到所需要的数据。

在用户搜索到需要的目标天体后,往往并不是直接下载下来;因为搜索到的数据量通常特别大,需要简单地预览一下搜索是否正确。这就需要数据库提供一种“预览”数据的功能。对于图像、光谱数据,可以通过图像直接展示,也可以输出其参数给用户参考。输出方式可以有多种形式:HTML 网页格式、CSV 格式、VO Table 格式^[27]或是直接的 FITS 文件等,最大程度地满足用户的需求。

4.3 数据库的下载服务

数据库最主要的服务就是让用户获取到数据。当用户寻找到需要的数据后,接下来就是如何把数据下载到自己的计算机或服务器上。一些零散的、少量的数据可以从数据库网页中直接下载。而一般来说,用户所需要下载的数据量都很大,如某个大天区所有目标的数据,直接从数据库网站页面访问下载不是很适合。这是因为由于 IO 带宽的限制,单个用户的大吞吐量需求很容易导致整个网路堵塞,使得其他用户不能正常地访问;或者是使得服务器搜索的时间过长,拖长整个查询下载过程的时间。对于这种情况,可以采取检索和下载服务使用不同网络通道的方法加以解决。下面以 NED 和 SDSS 的大量数据下载方法为例,说明如何实现这一过程。

NED 的作业批处理 (Batch Job) 服务流程为:(1) 用户以 E-mail 的形式向 NED 提交目标列表,这个目标列表按照 NED 网站的规定格式所填写,包括目标的基本信息,如目标名、位置坐标、搜索半径等参数;(2) 数据库服务器每 15 min 处理一次用户提交的申请,处理搜索结果,把搜索结果放到一个特定的地方,处理结束后,等待用户下载;(3) 用户点击“Pick

Up Batch Job Results” 页面, 以 FTP 方式下载数据, 如果数据量太大, 搜索花费时间会很长, 服务器系统也会通过 E-mail 的形式提醒用户下载数据。类似于这种数据下载方式的数据库还有 IRSA。除了以上的功能外, IRSA 还可以实时查询处理进度。

SDSS 的作业批处理 (MyDB 和 CasJobs^[28]) 服务流程为: (1) 用户在网站注册个人账户; (2) 登录账户, 在作业系统 (CasJobs) 网页内提交作业, 如果有多个用户同时申请或系统正忙, 进行作业排队, 等待完成; (3) 系统在用户的 MyDB 建立一个个人数据库, 在数据库系统里提取所需要数据下载到 MyDB; (4) 用户从 MyDB 的个人数据库系统下载 CSV, VO Table, FITS 等格式的数据到本机。

从以上的例子可以看出, NED 和 SDSS 的不同在于 NED 没有使用账户进行管理, 因此所有的用户公用共同的存储空间; 而 SDSS 引入了用户账号的概念, 可以为用户提供一个暂时的个人存储空间, 增加了私密性。虽然 NED 和 SDSS 在具体处理下载时的方式不同, 但它们的核心思想都是在下载批处理任务时, 将大量数据的搜索任务经后台处理后, 再单独开辟存储空间, 为用户另外开通网络通道提供下载。这样既可以避免大量的下载流量请求占用网页浏览通道, 又可以用户使用统一的渠道下载批量数据, 便于数据中心对数据下载进行管理。这两种下载方式, 都给天文学家带来了方便。

4.4 交叉认证服务

在两个不同的星表之间, 相同天体目标各自具有不同标识 ID 的现象普遍存在。在科研工作中, 往往期望针对同一目标在不同星表中获得联合搜索的信息, 于是就产生了不同星表间交叉认证工作的需求。通常, 交叉认证以目标源的位置为纽带, 将不同数据库中的数据联系起来, 从而获得多个数据库中的参数信息或多波段的数据信息。针对这种具体的交叉认证工作, 也有一整套的理论作为支撑来实现。

在交叉认证服务中, 莫过于 CDS 的 Simbad 和 NED 的名称解析服务最典型。它们分别维护着自己的一套数据库, 用户只需要提交天体目标的常见名称, 就可以获取目标的其它别名和代码。用户不仅可以登录 CDS 或 NED 的网站来直接利用它们的名称解析服务, 如果有需要, 还可以把名称解析服务的代码集成到自己程序中, 配合自己的程序完成交叉认证工作。

除了使用这两种名称解析服务进行交叉认证工作, 一些天文数据库网站还有一套符合自己需求的定制化交叉认证程序, 用于把站内多个数据库中的数据关联起来, 联合分析。CDS 的 X-Match 服务, 是一套典型的用于交叉认证工作的服务。它把自己站点内部的 Simbad 数据库、VizieR 数据库以及用户输入的源之间, 彼此或各自内部进行一对一、一对多、多对多等方式的交叉认证。MAST 下的 Cross-Mission^[11] 也是进行交叉认证工作的网上服务, 不过它的数据集范围只限在 NASA 系列的数据库。美国虚拟天文台 (VAO) 的 Cross-Comparison^[12] 是一款基于网页的在线交叉认证工具, 用户通过上传自己的目标源文件, 同 VAO 服务器提供的各个巡天项目的数据源相比较, 进而得出交叉认证结果。

4.5 多波段数据融合

随着科学技术的发展, 天文学进入了全波段巡天的观测阶段, 形成了多波段天文学。来

自各个波段的巡天和观测数据都在急剧增长,有了交叉认证工作后,这些星表就可以统一起来,以全方位地了解天体在各个波段的特性。通过多波段的交叉认证可以对天体的物理性质、演化规律获得更全面系统的认识,加深对认证目标源的理解,为统计分析和数据挖掘做准备。一般来说,多波段数据融合通常在一些综合类的数据库站点中提供服务,比如 MAST、HEASARC、IRSA、CDS 等。可以看出,这些含有丰富资源的数据库发布平台本身就包含了很多波段的数据,更适合将其所辖的各个波段的数据库加以整理,提供服务。

对于一些只针对某一项目的单一数据库,在其把自己的资源对外发布的同时,也可以集成同一检索条件的其他常用数据库资源的搜索结果,联合显示在结果下面,达到数据融合的目的。比如,当搜索的目标在自己的数据库存在相关资料时,可以优先显示自己数据库的内容,同时其他数据库的资源按照站点分类放在结果下面显示作为参考;当搜索的目标在自己数据库内没有资料时,在为用户提示没有找到数据的同时,也会调用其他数据库的资源展现给用户,实现“一搜多显”的一站式服务效果。这里可以结合实际情况,合理地利用其他数据库在本地的镜像资源。

多波段数据融合的另外一个应用在天文数据的可视化方面。类似于微软的 WWT 和谷歌的 Google Sky,都是把当前各个波段的图像数据(例如 DSS、2MASS、Chandra、Spitzer 等)进行处理,通过拼接、分层显示等方式实现最佳的可视化效果,再集成其深度挖掘出的各类信息,共同提供给使用者。事实上,我们通常所见到的很多漂亮的星系图片,也是利用各种不同波段的图片处理出来的。

4.6 其他服务

除了上面介绍的几个数据库的基本服务之外,实际应用中各个数据库的对外网站,还尽量加进一些针对自身数据的小的处理工具,使用户在线就能处理一些简单的问题。例如 NED 的网站就集成了许多实用的小工具:坐标转换和星际消光计算器、速率计算器、宇宙膨胀计算器等。另外类似于 ADS 和 SDSS 的网站,数据库站点会为用户提供一个用户账号,这样,用户的一些使用习惯可以记录在服务器上,用户常用的数据也可以缓存在自己的空间内,这在一定程度上可以更加个性化地为用户提供服务;账号的另外一个好处就是允许用户对具体数据做出自己的注释,方便自己和其他用户的理解。

像天文数据这样基于数据密集型的数据存储、管理、传输、处理模式,天文学家早就意识到有必要对所有的过程进行标准化管理。在这种背景下,一个跨天文、计算机和信息学科的概念——虚拟天文台(Virtual Observatory,以下简称 VO)诞生了。虚拟天文台是通过先进的信息技术将全球范围内的天文研究资源无缝透明连结在一起形成的数据密集型网络化天文研究和科普教育环境。虚拟天文台使数据的产生、存储、处理、发现和获取都能在标准的 VO 框架下进行。虚拟天文台的诞生,消除了各个数据库格式不统一的问题,使得数据交叉认证、图像光谱数据的分析等工作有了相应的 VO 工具来完成。通过使用这些工具,天文学家可以避免一部分重复性工作,节省了宝贵的时间。经过各国 VO 团队的努力,前面所述的一些服务,也已经部分或全部在 VO 的框架下有了具体的实现。如 VOSpec^[29]、Aladin、SPLAT^[30]、VOSesame^[31]、VOplot^[32]、TOPCAT^[33, 34]、Iris^[35]等都是 VO 出色的应用程序。目前,全世界已经有上百家天文数据中心或天文项目宣称支持 VO 的

标准, 为 VO 提供标准接口的数据源。可以想象, 如果未来的天文数据都在 VO 的标准下统一进行管理, 天文学家只需掌握 VO 的一些工具, 即可应用所有的天文数据来进行科研工作。

5 面向未来的天文数据管理与应用

从任何一个时期来看, 天文领域的发展都紧密依赖于当时的技术成果, 往往是当技术方面有大的提升后, 天文领域就有相应的突破。反过来, 天文科学研究尤其是天文海量数据的产生、存储和处理也深刻影响着相关技术领域的发展方向。

快速增长的海量天文数据对未来天文数据管理技术提出了新的需求。自计算机技术普及以来, 用于保存数据的存储设备发生了翻天覆地的变化。从早期的纸带、打孔卡、磁带、软盘, 到目前还在使用的 CD、DVD、U 盘、硬盘、存储阵列, 存储量成指数增长。伴随着计算机技术和信息技术的进步, 天文学在其中也受益匪浅。计算机技术的日新月异、元器件成本的不断下降等也使天文学家开展深层次巡天计划的信心不断增长, 一些超大口径超高效率的望远镜正在筹备建造, 随之而来的海量数据, 也给天文数据的管理带来了新的挑战。如前面所述的 LSST 项目, 每晚产出 TB 量级的科学数据, 这些数据都要在几个小时内完成存储、处理、归档等工作, 传统意义上的关系型数据库在处理上显得很吃力, 这就促进了新概念数据库的诞生。以 SciDB 为代表的新型数据库具有典型的非关系型科学数据库的特征。SciDB 的数据模型是基于多维阵列, 而不是一系列元组的集合, 这样就打破了传统数据库固定表结构的限制; 存储模型是基于版本控制而不是简单的替换更新, 这样在一定程度上能够提高数据库的操作速度; 支持内置的工作流模式; 面向网格的设计可以使其扩展到数千个节点的数百 PB 量级的数据存储量; 支持数据及其误差, 以适应科学领域一些数据的不精确性; 支持“外部的”数据对象, 以便数据集可以不需要加载到数据库才能被查询和操作……。总之, SciDB 是一个主要为科学领域中超大规模阵列数据而设计的开源数据管理和分析软件系统, 它代表了未来分布式数据库的一个发展方向。

未来的天文数据分享和交互, 需要更强大、更多渠道的网络互连通道来支撑。现代天文数据共享如此方便, 得益于深刻影响当今世界人们生活习惯的互联网。过去的 10 多年中, 互联网从无到有; 从最初的科研、军事部门专有, 到面向公众开放; 从几十 kb/s 的拨号上网, 到光纤通信进入普通家庭, 创造了一个又一个的神话。如今, 互联网移动时代已经到来, 各种终端的数据共享问题也摆到面前。网络硬盘、云存储等新的技术概念也正在逐步深入普通用户的思维中, 这也让我们依稀看到了一些天文数据存储的未来趋势。天文数据库以及天文数据管理在这场新的技术变革中, 应该有更大作为的空间。

虚拟天文台技术将在未来天文数据管理工作中充当重要角色。各个天文数据中心在发布、整理各个大天文数据集的同时, 也面临着如何科学地解决数据集之间数据融合的问题。事实上, 天文学家已经意识到, 在多数据集融合的环境下进行科研工作, 比在单个数据集下进行科研更具有优势, 能够提高科学发现的效率。未来的天文数据融合工作将更加深入。虚

拟天文台能够提供一个针对大型的复杂数据集进行互操作的环境,它把全球的天文数据资源加以整合,使其具有统一的规范和标准。目前,越来越多的数据库网站都加入了虚拟天文台技术,以适应未来数据服务的需求。

新的自主观测方式也丰富了天文数据的管理模式。技术的进步和天文学家的需求催生出更先进的观测模式——程控自主天文台^[36]。这是一种完全没有人干预的观测方式,从观测计划的制定、任务队列的调度、望远镜系统的控制、CCD 相机的曝光、数据存储处理和发布都是自主化的过程。事实上程控自主天文台从概念到实验,也经历了几十年的时间^[37-43],直到最近的 10 年,才出现几个真正意义的自主天文台^[44-48]。由于全部过程自动化,只需要预先设计好数据库的格式和存储规范,天文学家就可以完全实时地得到天文数据。这对天文学家来说,无疑是最期望的结果。

PB 甚至 EB 量级天文数据的产生、存储、传输、检索和计算等强烈地依赖于不断发展的硬件和软件技术。天文学家连同信息、计算机领域的专家正为解决海量天文数据存储、传输、处理等问题而努力。一些最新的高效技术如并行存储、GPU 计算、非关系型数据库、可视化分析等正逐步地应用到天文领域中,再结合 IVOA 制定的虚拟天文台框架协议,可以实现一个整套的天文数据应用解决方案。未来的天文数据应用也许只是用户拿着移动终端,登陆自己的“天文空间”就能以可视化的方式完成必要的科研工作;而这个“天文空间”是已经在虚拟天文台框架下高度整合好的,集合所有天文数据的大数据集,用户无需关心后台的任何技术细节,并且可以按照每个用户的使用习惯定制化需求;用户还可以在线查看自己向某一个观测站所提交的观测计划的观测进度等信息。

6 结 语

在天文数据库系统方面,我国起步较晚;然而在国内天文学家的努力下,逐步掌握和吸收了近年国际上优秀天文数据库技术。2002 年,中国虚拟天文台团队 (China-VO) 成立,从此,在实现天文数据库技术方面,国内有了专业的组织和渠道。在国家大科学装置郭守敬望远镜 (LAMOST) 项目的支持下,China-VO 开发出了一系列实用的工具,如 SkyMouse、VOTFilter、FitHas、VO-DAS、FITS Manager 等^[49]。当前,LAMOST 的先导巡天工作已经结束,先导巡天数据的发布系统就主要参考了各国天文数据库的优秀服务以及 IVOA 制定的标准:数据发布网站汇集了直接浏览天区检索、锥形检索以及针对 LAMOST 数据参数检索等多种检索方式,并集成了 CDS 的 sesame 名称解析服务;在数据下载服务方面,用户可以采取浏览单个目标数据后直接下载,或是结合每个用户的标识码进行 wget 方式的批量下载两种方式;数据的输出除了 VOTable、FITS、CSV 格式外,还可以把选定目标在 Aladin applet 上显示出来,通过叠加国际上其它的星表进而达到和其它数据库数据融合的目的^[50]。目前,China-VO 正努力使 LAMOST 的数据库服务支持更多的 IVOA 协议和标准,使 LAMOST 的数据用户有更加舒适的使用体验。

在 China-VO 研发团队的努力下,中国天文数据中心^[51] 新的网站于 2011 年正式对外

开放。中国天文数据中心是中国科学院国家天文台为天文学的科学研究、普及教育以及相关学科和社会需求提供天文科学数据服务的基础设施, 它在原挂靠在中国科学院国家天文台(原北京天文台)的世界数据中心天文学科中心的基础上发展而来。目前中国天文数据中心包含的资源有郭守敬望远镜科学试观测与先导巡天数据、BATC 大视场多色巡天观测数据、CSTAR 中国之星测光数据、南银冠 U 波段巡天数据、2.16 m 望远镜观测数据, 以及 SDSS DR8、2MASS、GSC 2.3、UCAC3、CDS ViZieR、NASA ADS 镜像数据等。随着自身实力的不断增长, China-VO 团队正以中国天文数据中心为平台, 努力为国内外天文用户提供多方面高质量的天文数据服务。

我们正步入天文数据 PB 量级的时代, EB 量级时代也不再是遥不可及, 如此大数据集和数据流的有效管理确实面临着巨大的挑战。然而从整个天文发展历史来看, 数据存储与管理工作的意义, 就是为未来的科学研究积累宝贵的历史财富。天文数据库服务是典型的数据密集型服务。现代天文的迅速发展, 得益于先进的信息和计算机技术。天文学研究的方式在科技的发展下已经改变了很多, 而这种趋势仍然在继续。

致谢

感谢 LAMOST 大科学工程提供的良好工作环境, 感谢中国虚拟天文台团队的大力支持。

参考文献:

- [1] Pence W D, Chiappetti L C, Page C G, et al. *A&A*, 2010, 524: A42
- [2] Djorgovski S G, Mahabal A A, Drake A J, et al. <http://arxiv.org/ftp/arxiv/papers/1203/1203.5111.pdf>, astro-ph/1203.5111, 2012
- [3] Rawlings S, Schilizzi R. <http://arxiv.org/pdf/1105.5953v1.pdf>, 2012
- [4] LSST — Large Synoptic Surver Telescope, <http://www.lsst.org/lst/>, 2012
- [5] Kaiser N, Aussel H, Boesgaard H, et al. *Proc. SPIE*, 2002, 4836: 154
- [6] Brown R L, Wild W, Cunningham C. *Advances in Space Research*, 2004, 34: 555
- [7] Pesnell W D, Thompson B J, Chamberlin P C. *Solar Physics*, 2012, 275(1-2): 3
- [8] Keil S L, Rimmele T R, Wagner J, et al. *Astronomische Nachrichten*, 2010, 331(6): 609
- [9] Williams R, Hanisch R, Szalay A, et al. <http://arxiv.org/ftp/arxiv/papers/1110/1110.0498.pdf>, astro-ph/1110.0498, 2012
- [10] CDS X-Match Service, <http://cdsxmatch.u-strasbg.fr/xmatch>, 2012
- [11] MAST Cross-Mission, <http://archive.stsci.edu/xcorr.php>, 2012
- [12] VAO Cross-Comparison, <http://vao-web.ipac.caltech.edu/applications/VAOSCC/>?, 2012
- [13] VOSpec: a VO Spectral Analysis Tool, <http://www.sciops.esa.int/index.php?project=ESAVO&page=vospec>, 2012
- [14] The Aladin Sky Atlas, <http://aladin.u-strasbg.fr/aladin.gml>, 2012
- [15] Strasbourg Astronomical Data Center (CDS), <http://cds.u-strasbg.fr/>, 2012
- [16] SIMBAD Astronomical Database, <http://simbad.u-strasbg.fr/simbad/>, 2012
- [17] VizieR Service. <http://vizier.u-strasbg.fr/viz-bin/VizieR>, 2012
- [18] The Sloan Digital Sky Survey, <http://www.sdss.org/>, 2012

- [19] Szalay A, Gray J, Thakar A R, et al. Microsoft Research Tech.Rep, 2001: 104
- [20] ESO Archive Query Form, http://archive.eso.org/eso/eso_archive_main.html, 2012
- [21] CADDC - The Canadian astronomy Data Centre, <http://cadcwww.dao.nrc.ca/cadc/>, 2012
- [22] Szalay A, Gray J. *Science*, 2001, 293: 2037
- [23] Abell P A, Allison J, Anderson S F, et al. <http://arxiv.org/pdf/0912.0201v1.pdf>, 2012
- [24] Strauss M A, Tyson J A, Anderson S F, et al. <http://arxiv.org/pdf/0903.3149v1.pdf>, 2012
- [25] Valdúriez P. *Int. Conf. on Databases and Expert Systems applications*, 2011, 6860: 1
- [26] Malon D, Cranshaw J, Gemmeren P, et al. *Journal of Physics: Conference Series, International Conference on Computing in High Energy and Nuclear Physics*, 2001, 331: 042016
- [27] Ochsenbein F, Williams R, Davenhall C, et al. 2011, preprint(astro-ph/1110.0524)
- [28] O'Mullane W, Li N, Nieto-Santisteban M, et al. Microsoft Research Tech. Rep, 2005: 19
- [29] Osuna P, Barbarisi I, Salgado J, et al. *Astronomical Society of the Pacific Conference Series, Astronomical Data Analysis Software and Systems XIV*, 2005, 347: 198
- [30] SPLAT - Spectral Analysis Tool. <http://star-www.dur.ac.uk/pdraper/splat/splat.html>, 2012
- [31] VOSesame. <http://iraf-nvo.noao.edu/vo-cli/vosesame.html>, 2009
- [32] Kale S, Vijayaraman T M, Kembhavi A, et al. *Astronomical Society of the Pacific Conference Series, Astronomical Data Analysis Software and Systems XIII*, 2004, 314: 350
- [33] Miller C J. In: Graham M J, Fitzpatrick M J, McGlynn T A, eds. *The National Virtual Observatory Book ASP Conference Series*, 2007, 328: 19
- [34] Taylor M B. In: Shopbell P L, Britton M C, Ebert R, eds. *Astronomical Data Analysis Software and systems XIV ASP Conference Series*, 2005, 347: 29
- [35] Doe S, Bonaventura N, Busko I, et al. <http://arxiv.org/pdf/1205.2419v2.pdf>, 2012
- [36] Castro-Tirado A J. *Advances in Astronomy*, 2010, 570489: 1
- [37] McNall J F, Miedaner J L. *Astronomical Journal*, 1968, 73: 756
- [38] Lee R H, MacQueen R M, Mankin W G. *Applied Optics*, 1970, 12(9): 2653
- [39] Dennison E W. *Proceedings of Conference on Large Telescope Design, ESO/CERN Conference, Switzerland: Geneva*, 1971: 363
- [40] Robinson L B. *Annual Review of Astronomy and Astrophysics*, 1975, 13: 165
- [41] Skillman D. *Sky & Telescope*, 1981, 61: 71
- [42] Genet R M. *International Amateur-Professional Photoelectric Photometry Communication*, 1986, 25: 1
- [43] Taki T. *Sky & Telescope*, 1989, 77: 194
- [44] Vestrand W T, K. Borozdin N, Brumby S P, et al. *Advanced Global Communications Technologies for Astronomy II, of Proceedings of the SPIE*, 2002, 4845: 126
- [45] Zerbi F M, Chincarini G, Ghisellini G, et al. *Instrument Design and Performance for Optical/Infrared Ground-Based Telescopes, of Proceedings of the SPIE*, 2002, 4841: 737
- [46] Hessman F. <http://www.astro.physik.uni-goettingen.de/~hessman/MONET/links.html>, 2012
- [47] Lipunov V M, Kornilo V G, Krylov A V, et al. *Astrophysics*, 2005, 48(3): 389
- [48] Baruch J. *Astronomy & Geophysics*, 2007, 48, 4: 27
- [49] China-VO services. <http://www.china-vo.org/service.html>, 2012
- [50] LAMOST Pilot Sky Survey Data. <http://lamost.escience.cn/dct/page/66293>, 2012
- [51] CAsDC. <http://casdc.china-vo.org/>, 2012

Review and Prospect of the Astronomical Database

LI Jian, CUI Chen-zhou, HE Bo-liang, ZHAO Yong-heng, CAO Zi-huang,
FAN Dong-wei, LI Chang-hua, CHEN Yue

(*Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100012, China*)

Abstract: Modern astronomical research relies heavily on the scientific databases. In the past few decades, the advance of technology, especially that in the computer and information sciences, brought revolutionary change in astronomical database. Nowadays, the routine of astronomical research usually is as follows: logging to the astronomical database website, entering the key words to query, downloading astronomical data, and then analyzing the data set.

The early prototype of astronomical database was established in 1970's. Strasbourg Astronomical Data Center (CDS) is the pioneer in the astronomical database field. Simbad service in CDS is the earliest scheme providing astronomers with a convenient method to query against astronomical objects. After late 1990's, the database has been widely used in various fields including astronomy. With the rapid development in information technology, nowadays, the astronomical data can be accessed by astronomers more easily. A great number of astronomical database/service websites were established for example: NED, MAST, ADS, SDSS, 2MASS. These databases provide 7×24 hours continuous data services to astronomers.

Simple data searching and downloading are only the most basic services of these database websites. Besides data access, other services are also provided. Cross-identifying sources among multiple catalogues is a general method to investigate an object in astronomical research, which is especially important for the data fusion across multi-wavelength astronomy. Astronomical object name resolver service is another useful tools, which is usually combined to the search engine for giving a more convenient data searching service. Recent years, data mining and other data processing mechanism based on the astronomical databases become more and more popular. The databases can be used for a much broader variety of studies than it was possible in the past.

With the coming of the large sky survey and time domain astronomy period, the scale of astronomical data grows exponentially. There are challenges in managing and mining these large and complex data sets. Powerful computing hardware, networks and data management system are required for more efficient data storage, processing, access, visualization and analysis. Virtual Observatory (VO) was proposed at the beginning of 21st century to meet

these astroinformatic challenges. Now there is an increasing number of astronomical database websites that are VO-compliant. In this data-intensive science era and under the frame work of VO, research in astroinformatics will supply more useful tools for astronomers.

Key words: astronomical database; data query; cross-match; sky survey; Virtual Observatory

《天文学进展》2013 年征订启事

《天文学进展》是天文学类中文核心期刊。刊物为季刊，2013 年本刊于 2、5、8、11 月中旬出版，每期定价 40 元，全年 160 元（含邮寄费和包装费）。凡需订阅 2013 年《天文学进展》，请到所在地邮局进行订购。

统一刊号：CN 31-1340/P

邮发代号：4-819

欢迎订阅！谢谢支持！

《天文学进展》编辑部