

互联网时代的天文学革命----

虚拟天文台

赵永恒

(国家天文台 北京 100012)

四百年前伽利略发明了天文望远镜,使人类首次摆脱了用肉眼直接观测天空的历史,为从哥白尼开始的天文学革命提供了大量的科学证据。历史悠久的天文学经过哥白尼、伽利略、开普勒和牛顿等人的发展,演变成了一门崭新的科学,同时也催生了现代科学技术。

到一百五十年前,由于照相技术和光谱技术在天文观测中的应用,用人眼作为天文探测器的时代也结束了,由此诞生了天文学的新分支——天体物理学,并发展成为现代天文学的主流。

五十多年前,在第二次世界大战中得到蓬勃发展的无线电技术使得人类的视野跃出了可见光的波段,发展成了射电天文学。之后不久随着宇航时代的到来,天文观测不再局限于地面,由此诞生了空间天文学,人类对宇宙的观测扩展到了伽马射线、X射线、紫外和红外波段。

从十多年前开始,天文学正在进行着革命性的变化,这一变化是由前所未有的技术进步所推动的,即望远镜的设计和制造、大尺寸探测器阵列的开发、计算能力的指数增长以及互联网络的飞速发展。

互联网时代的天文学

望远镜技术的进步使得人类可以建造大型的空间天文台,为伽玛射线、X射线、光学和红外天文的发展开辟了新的前景,同时也推动了新一代的大口径地面光学望远镜和射电望远镜的建造。在光学与近红外波段,已经有了高灵敏度高分辨率而尺寸不断增大的探测器阵列。伴随着这些技术进步,天文学家们正在计划建造功能更好口径更大的空间和地面望远镜,并将配备尺寸更大像素更多的探测器。如同在计算机工业中反映计算能力随时间指数增长的摩尔定律一样,在过去十年中技术进步使得天文学的发展实际上也是遵循摩尔定律的^[1](图1)。

随着众多先进的地面与空间天文设备的投入使用,将产生大规模的观测数据,例如目前哈勃空间望远镜每天大约产生50亿字节的数据,我国正在建造的LAMOST望远镜也将产生每天30亿字节的数据,而美国计划建造的“大口径巡天望远镜”将会达到每天10万亿字节的量级!

除了数据量的快速增长外,天文观测的方式也有了变化。这些先进的地面和空间天文台更多的是进行大规模的巡天观测,并产生质量均匀、标准统一的海

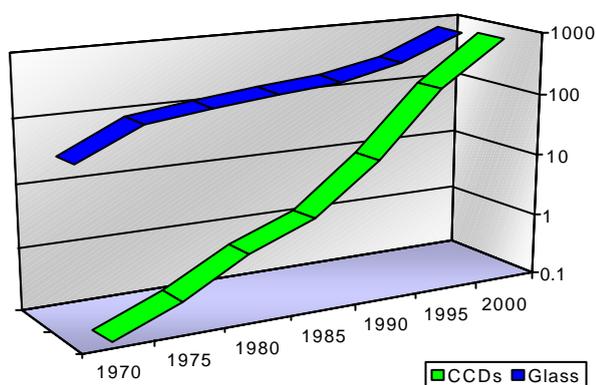


图1 光学望远镜接收面积(平方米)和 CCD 像素(百万像素)的增长曲线。望远镜面积每 25 年增长一倍,而 CCD 像素每 2 年增长一倍。

量数据，通常是万亿字节的量级。这种天文观测模式的变化，不仅由于这些新设备能够快速获得观测数据，还由于计算机软硬件技术的发展允许对数据进行快速的采集、处理及存档。

每个这样的巡天数据库从其本身来说都是非常有价值的，但是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，如蟹状星云（图 2）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 neutron 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。

使天文学研究发生变化的另一个主要的技术进步是宽带互联网络技术，它允许在不同地点间进行天文数据的交换，使世界各地的天文学家都能够访问和使用这些数据，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：

1. 天文数据绝大部分是开放数据。天文界的传统是所有的观测数据在一年后向公众开放，使得观测者有时间进行数据分析和发表早期结果，也使其他天文学家可以有机会使用这些数据。国际上的许多大型天文观测项目的观测数据都会及时在互联网上公布，这为数据共享提供了良好的基础，这在各类学科中是独一无二的。

2. 天文数据的数据量非常大。现有的以及即将实施的天文项目每天都会产生数十亿字节甚至上千亿字节的数据，天文数据中心的存贮容量已经达到数千亿字节，并开始向千万亿字节扩展。

3. 天文数据有比较好的归档，并提供互联网服务。当前，世界上已经有多家天文数据中心在天文数据归档方面做了大量的工作，并取得了很好的应用。甚至各类天文文献也是在线网络服务的。

4. 天文数据的格式多种多样。天文数据的内容主要有星表、星图、光谱等，数据的格式则各种各样，其内部格式都依不同的天文观测项目而变化。

5. 天文数据是全波段的数据。从伽玛射线、X 射线、紫外、光学、红外到射电波段都有观测项目在进行，这些数据是高度相关的，需要在高维参数空间内进行研究。

6. 新的观测数据总能带来全新的现象或规律的发现。

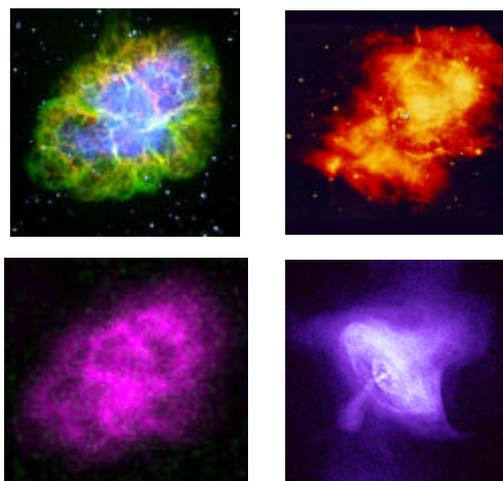


图 2 蟹状星云的多波段图象（左上：光学、左下：射电、右上：红外、右下：X 射线）。蟹状星云是 1054 年中国天文学家发现的一颗超新星爆发后所留下的遗迹。

虚拟天文台

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，因此，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^[2]：

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) 数据挖掘技术：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^[3]。

虚拟天文台与网格技术

和传统概念上的天文台不同，虚拟天文台是以高速发展的互联网技术和计算机科学为支

撑的，因而具有其自己的鲜明特点。首先，虚拟天文台是快速发展的，随着数据量的迅速增加、计算机网络的快速发展，虚拟天文台的各种软硬件设施都要及时更新以保持其先进性；第二，虚拟天文台是分布式的，无论其数据还是各种计算机软硬件资源都将分布在不同的国家和地区；第三，虽然在物理构成上具有分布性，但在功能上必须是有机的统一体，从而为用户提供整体一致的服务；第四，虚拟天文台是面向全球的，它的资源将为全世界的天文学家所共享。此外，拥有独一无二的科学资源的虚拟天文台必将承担起与它相称的社会责任，为公众了解和利用科学提供服务。

虚拟天文台主要由以下几部分组成（图3）：1）数据档案，数据档案里存储着数据集（如星表、图像和光谱），组织成逻辑相关的数据集合，也存储着描述数据档案及其数据存储方式的元数据。用户可以通过各种方式进行网络访问。2）元数据标准，元数据是描述虚拟天文台的一些要素的结构化信息，用来描述档案和档案所提供的服务。3）数据访问层，数据访问层将提供统一的界面来访问虚拟天文台中所有的数据、元数据和计算服务。4）分析工具，当数据访问层和元数据标准允许虚拟天文台连接档案和访问数据时，就需要分析工具来支持信息发现和提供数据挖掘所需要的统计相关和图像分析的能力。5）数据挖掘工具，包括大型的多变量数据集的可视化和统计分析工具，让用户在极其丰富的数据海洋里挖掘出新的成果。

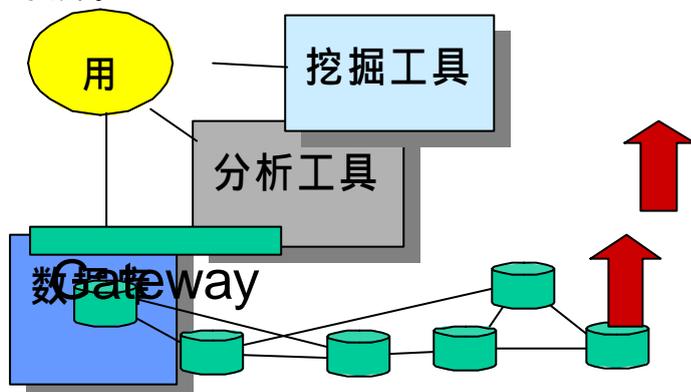


图3 虚拟天文台的体系结构

虚拟天文台这样的体系结构与第三代互联网技术即网格（Grid）技术是异曲同工的。

第一代互联网是在二十世纪七八十年代将分布在世界各地的计算机用 TCP/IP 协议连接起来，主要的应用是电子邮件。第二代互联网是二十世纪九十年代将成千上万网站上的网页连接起来，主要的应用是网页信息浏览以及电子商务等

信息服务。而第三代互联网技术就是网格技术^[4-6]。

网格技术要解决的信息共享不是一般的文件交换与信息浏览，而是要把所有网络单位连接成一个虚拟的社会组织，实现在动态变化环境中具有灵活控制的协作式信息资源共享。现有的网页服务器就好像互联网世界上一个个孤立的小岛，虽然这些“孤岛”之间暂时还有充足的带宽资源可用，但大量的信息还是被“锁”在各个小岛的中央数据库里，各“孤岛”之间并不能按照用户的指令进行有意义的交流。而网格技术是要利用现有的网络基础设施、协议规范、互联网和数据库技术来为用户提供一体化的智能信息平台，其目标是创建一种基于互联网的新一代信息平台 and 软件基础设施。在这个平台上，信息的处理是分布式、协作和智能化的，用户可以通过单一入口访问所有信息。网格追求的最终目标是能够做到服务点播和一步到位的服务。简单地讲，网格是把整个互联网整合成一台巨大的超级计算机，实现计算资源、存储资源、数据资源、信息资源、知识资源、专家资源的全面共享。

网格系统可以分为五个基本层次：1）构造层，它的功能是向上提供网格中可供共享的资源，它们是物理或逻辑实体。2）连接层，它是网格中网络事务处理通信与授权控制的核心协议。构造层提交的各种资源间的数据交换都在这一层的控制下实现。各资源间的授权验证、安全控制也在这里实现。3）资源层，这一层的作用是对单个资源实施控制，与可用资源进行安全连接、对资源做初始化、监测资源运行状况、统计与付费有关的资源使用数据。4）汇集层，这层的作用是将资源层提交的受控资源汇集在一起，供虚拟组织的应用程序共

享、调用。5) 应用层。这层是网络上用户的应用程序。应用程序通过各层的 API 调用相应的服务，再通过服务调用网络上的资源来完成任务。

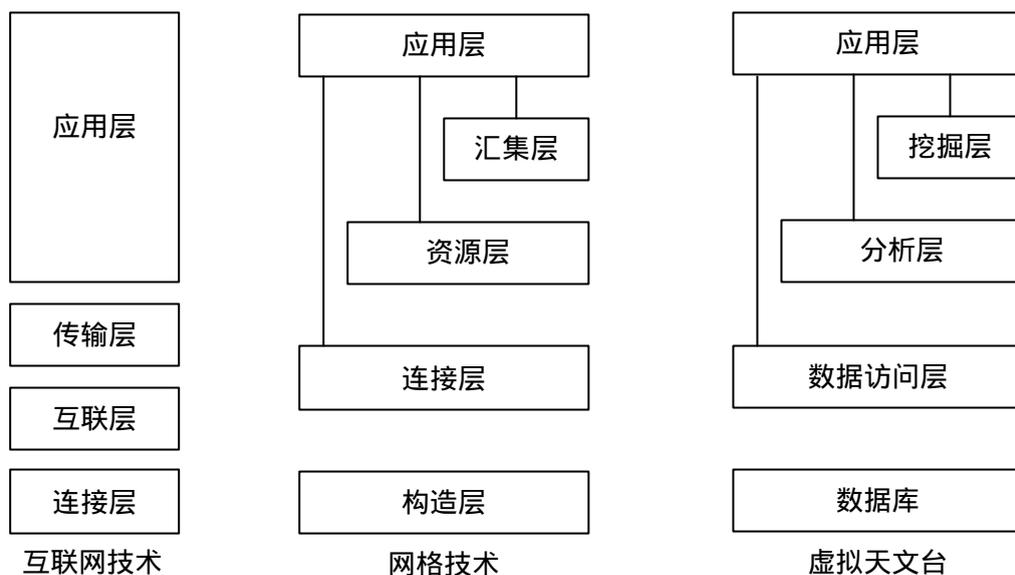


图 4 互联网技术、网络技术和虚拟天文台的对应关系

虚拟天文台实际上是数据网格与信息服务网格的综和（图 4），在一定程度上还包括计算网格的功能。因此虚拟天文台的发展目标为网格技术提供了独一无二的试验场，从网格基础设施的构建，到网格操作系统的开发，最后到网格天文应用工具的实现，虚拟天文台为网格技术提供了一整套的应用需求。

虚拟天文台所面临的技术挑战是十分明显的：星表的大小是万亿字节，而各类数据的总量将达到千万亿字节；这些数据是非常复杂的，包含上亿甚至更多的天体，每个天体又有几十或几百个属性，这对于数据挖掘而言是一个至关重要的新问题；而对如此庞大的星表进行多变量相互关联将是一个大规模计算的问题；如果还要对天体进行像素层次上的分析，计算的问题将更加突出。此外，还需要对元数据的表示与处理、大规模统计分析与相互关联、以及分布式并行计算技术等进行多学科研究，来解决虚拟天文台所面临的前所未有的数据访问和计算问题。

实际上，不单单是天文学面临这个问题，诸如高能物理、计算基因学、全球气候研究和海洋学等其它科学分支也和虚拟天文台一样面临着类似的技术挑战。例如，人类基因组数据大小约为 30 亿字节，而整个天空的数字巡天的数据量约为 10 万亿字节，处理这样量级的数据的工具和技术明显地需要发展新的信息技术来予以支持。当前，信息技术的研究与开发已在一些领域中展开，如大型数据库的统计分析与数据挖掘、分布式高性能计算网格、数据增强网格计算（数据网格）以及结构化数字信息管理（数字图书馆），这些研究在许多方面都和虚拟天文台所面临的问题有关，贯穿于这些科学分支中的信息技术和数据管理技术将会推动虚拟天文台的发展。

可以预计，虚拟天文台将在存储技术、信息管理、数据处理、分布和并行计算、高速网络、数据可视化和数据挖掘等各个领域突破现有技术的限制。这就要求天文界和信息科学界携起手来共同研究和开发虚拟天文台所需要的信息技术，要求与其它学科和与超级计算中心进行合作，来开发元数据处理、数据处理和分布式计算的标准。而数据挖掘是一个需要天文学家、计算机科学家、数学家和软件专家一起合作才能解决的多学科问题。

大规模的数据以及用户与资源的广泛分布同样对网络的互联性提出巨大挑战，虚拟天

文台将是利用广域的高性能宽带网络来进行学术研究的一个富有创造性的例子。

因此，虚拟天文台的实现不可能仅是天文学家的事情，必须与计算机、网络、统计学等领域的专家共同努力，这无论对天文学还是对信息科学等都是双赢甚至多赢的合作。

中国天文学的发展机遇

1999年美国天文学家首先提出了虚拟天文台的概念，很快就在国际天文界引起了广泛的关注。美国国家科学院天文学及天体物理学发展规划委员会在题为“新千年的天文学和天体物理学”的十年发展规划中把建立美国国家虚拟天文台（NVO）作为优先推荐项目，目前已通过美国国家科学基金立项，获得一千万美元的资金资助，历时五年。

欧洲南方天文台（ESO）、法国斯特拉斯堡天文数据中心（CDS）、空间望远镜欧洲协调机构 ST-ECF、巴黎天体物理研究所和英国曼切斯特大学等联合提出了欧洲天体物理虚拟天文台（AVO）的项目，如今已获批准并得到近三百万英镑的启动资金。

英国天文界为了保持自己的领先优势，提出了天文网格（AstroGrid）的项目，获得资金五百万英镑，并成为英国科学界 e-Science 项目的重要组成部分。

加拿大天文数据中心决定发展从大型科学数据库中进行数据挖掘的工具，提出了发展数据挖掘技术的提案，该项目历时三年，直接费用一百五十万美元。

我国天文学家对虚拟天文台也有很高的热情并积极参加到建设中去，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”，以研究和跟踪国际上相关领域的发展动向、确定“中国虚拟天文台”的建设目标、推动国内天文界和相关学科对虚拟天文台的研究与发展。目前，国家天文台、北京大学、清华大学等单位正在积极有效地推动相关的工作。

虚拟天文台的出现为中国天文学提供了新的发展机遇，这是因为：1）与在信息技术领域中互联网的建设相似，中国天文界只有在虚拟天文台的建设中做出自己的贡献，才有可能充分利用虚拟天文台来获得科学上的发展；2）目前我国天文界业已开展了一些小规模的天文项目（如BATC巡天、密云米波射电巡天和超新星巡天等），而在未来随着我国天文界的重大科学工程项目如LAMOST望远镜、太阳空间望远镜（SST）和五百米口径射电望远镜（FAST）等一批大型观测设备的建成，中国天文学将在国际天文学界拥有更高的发言权，并将在未来的虚拟天文台中发挥重要的作用；同时，这些大型观测设备的科学产出也需要虚拟天文台这样的机构来配合；3）虚拟天文台对于我国这样的发展中国家来说具有更深刻的意义，国力的限制使得我们不可能全面地发展各类型的大型天文观测设备，而由于国际上天文观测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的数据来做出一流的科学研究工作；4）由于虚拟天文台为天文学开辟出了一条崭新的研究道路，因此必须在我国培养新一代的天文学家以适应时代发展的需要；5）虚拟天文台的设想的提出仅有三年的时间，可以说我们是与其它国家处于同一起跑线上，而我国飞速发展的信息技术和计算机网络技术也提供了强有力的保障；6）虚拟天文台的出现为天文学家、计算机科学家、数学家和统计学家的精诚合作提供了良好的机遇，从而为我国多学科跨领域的交叉研究提供了新的舞台。

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、促进新发现的催化剂，它将开辟出探索宇宙的新天地。在这历史性的机遇面前，中国科学家们应该也一定能够为虚拟天文台的建设与发展做出自己应有的贡献。

参考文献

- [1] Szalay A., Gray J. The World-Wide Telescope, *Science*, 2001, **293**:203
- [2] Brunner R., Djorgovski S., Szalay A. Towards a National Virtual Observatory, *Virtual Observatory of the Future*, Michigan, 2001, p.343-372
- [3] Banaday A., Zaroubi S, Bartelman M. Mining the Sky, Springer-Verlag, Berlin, 2001
- [4] 李国杰 信息服务网格—第三代 Internet, *计算机世界*, 2001, **40**:B8
- [5] 李国杰 超级服务器与信息网格, *科学新闻周刊*, 2001, 42:6
- [6] 徐志伟等 网格技术专题 *计算机世界*, 2001, **43**:B1-B9

关键词 虚拟天文台 网格技术 信息科学

Revolution of Astronomy on Internet—**Virtual Observatory**

Zhao Yongheng

(National Astronomical Observatories, CAS, Beijing 100012)

Abstract

With the developments of the astronomical techniques and information sciences, it is the time of astronomy with an enormously increased flow of data. This paper describes the function of the Virtual Observatory for the revolution in astronomy, and its relationship with grid techniques. The Virtual Observatory will provide new opportunities for Chinese astronomical community.

赵永恒 男, 国家天文台研究员、博士生导师。1964年2月出生。1990年获南京大学理学博士学位。国家“九五”重大科学工程LAMOST项目常务副总经理, 世界数据中心中国天文学科中心主任。主要从事活动天体的理论研究、高能天体的观测分析、多波段观测等研究工作, 参加和主持过“攀登计划”项目、国家自然科学基金项目等研究工作。1997年入选中国科学院“百人计划”, 1998年获中国科学院第三届十大杰出青年称号, 1999年获教育部科技进步奖三等奖。发表学术论文50余篇。