

文章编号: 1000-8349(2006)02-0093-07

虚拟太阳天文台及其发展

林钢华

(中国科学院 国家天文台, 北京 100012)

摘要: 伴随着太阳物理研究的深入、观测仪器的发展、观测数据的积累, 促使人们去思考这样一个问题: 如何使研究者能够方便地检索、分析和使用在不同时间、不同地点、不同波段上观测得到的大量与太阳有关的数据, 从而能探索更多悬而未决的科学谜题。这就是虚拟太阳天文台 (VSO) 项目被提出后得到有关天文台、研究所和大学积极响应并迅速投入运作的原因所在。介绍了虚拟太阳天文台的由来、作用、采用的技术和发展状况等。

关键词: 天文学; 虚拟太阳天文台; 综述; 太阳物理; 数据档案; 联机检索工具; 网络; 网格
中图分类号: P1 ; N37 ; TP39 **文献标识码:** A

1 虚拟太阳天文台简介

1.1 定 义

虚拟太阳天文台 (Virtual Solar Observatory, VSO) 是一套分布式太阳档案和分析软件的联合系统, 通过它能够实现对世界上不同地点天文台拥有的太阳数据的查询、获取和分析。由于这个系统具有提供太阳的天文学信息的“天文台”功能, 但其以大量太阳信息的形式存在于 Internet 上, 并不是一个物理实体, 所以称其为虚拟太阳天文台。

1.2 概念的提出

目前全世界有 60 多个太阳天文台, 有的拥有综合性仪器可获取一系列标准波长范围的每日全日面图像, 有的专门获取物理参数 (如速度场和磁场), 有的则以很高的空间和时间分辨率去研究太阳上的某些特定现象, 如活动区或日珥。然而, 无论这些数据档案在局域范围里多么完整, 都不能与其他类似档案直接相连。为了寻找匹配专门标准的数据, 研究人员不得不访问很多单个档案站点, 联系很多数据负责人。Roudier 和 Malherbe^[1] 指出, 1995 年仅在欧洲就有几十个可用的数据库。显然, 访问这些数据库搜索有关的数据是非常耗时的任务, 甚至几乎不可能完成^[2]。另一方面, 研究人员搜索少数几个著名天文台的观测数据时, 通常得到的也只是极其简化的数据, 不能满足科学研究的需要。

收稿日期: 2005-02-16 ; 修回日期: 2006-01-25

基金项目: 国家自然科学基金重点基金资助项目 (90412016)

运行在哥达得的太阳数据分析中心是试图把太阳数据集中在一个单一位置的例子, 但它的数据库只局限于主要天文台和卫星。1997年 Sanchez-Duarte 等人^[3]首先提出将太阳档案综合成一个统一系统的思想。1998年 Dimitoglou 等人^[2]进一步讨论了这一思想。同年, 意大利 Capodimonte 天文台的 Kevin Reardon 提出^[4], 太阳活动周期变化的详细研究需要分析横跨多年的观测记录, 这需要根据标准协议将多个单独维护的数据库无缝隙地连接起来, 实现跨多个档案的复杂检索。意大利 8 个天文台和研究所的天文工作者也提出“全部太阳数据目录”(Whole Sun Catalog, WSC) 项目^[5], 即有效连接遍布世界各地的可用的已提供的单个太阳档案, 简化执行检索任务和扩展研究人员的查询范围。

1998年 Hill^[9]详细描述了 VSO 的概念。2001年 7 月 NASA 空间科学日地结合部举行了两年一次的运作使命和数据中心回顾会议^[6], 其中太阳数据分析中心 (SDAC) 也被包括在此次会议内。会上对 VSO 提出了研究、原型、开发方面的经费预算。在此基础上, 2002年 11 月由美国斯坦福大学、国家太阳天文台、蒙大纳州立大学和太阳数据中心组成的 VSO 研究组发布了《虚拟太阳天文台设计协议书》^[7], 给出了 VSO 的体系结构、组织结构、特性、技术方法、首批参加此项目的成员、经费分配、项目运作时间表等。至此 VSO 项目正式启动。由前期发展而演化为 VSO 的项目还有 JOSO (Joint Organization for Solar Observation)^[8]、SOLAR (SOHO Long-term Archive)^[9]、ARTHEMIS (Archive Thermal Emission Imaging System)^[9]、法国基于地基太阳观测的长期太阳档案项目 BASS2000^[10]等。与此同时, 数据档案及其挖掘对于太阳研究的重要性, 得到了美国国家研究委员会空间研究理事委员会的认可^[11]。1998年该委员会推荐“开发一个由 NSF (National Science Foundation) 和 NASA (National Aeronautics and Space Administration) 联合支持、通过 WWW 访问的分布数据档案”。嗣后, 美国国家研究委员会天文学天体物理调查委员会郑重签署了有关协议, 把国家虚拟天文台 (National Virtual Observatories, NVO) 作为顶级优先启动的小型项目。NVO 是一个与 VSO 相似的项目, 它显示了天文界对交叉相关可行性研究日益增长的关注。

1.3 数据组成

VSO 得到了著名太阳数据中心的积极响应, 最初就有 10 个单位计划提供各自拥有的数据作为 VSO 的一部分, 这些数据源于地面和空间观测, 主要来自美国 (包括 NSF 和 NASA 基金组), 当前数据总量为 104.3 TB。原先主要由 NSO (National Solar Observatory)、SDAC (Solar Data Archive Center)、斯坦福大学和蒙大纳州立大学提供数据服务。NSO 提供的数据库包括 GONG (Global Oscillation Network Group)、GONG+、SOLIS (Synoptic Optical Long-term Investigations of the Sun) 等; SDAC 提供各种空间和补充数据; 斯坦福大学提供 Wilcox 太阳天文台、SOHO MDI (Michelson Doppler Imager) 及其他日震观测数据源; 蒙大纳州立大学则提供扩展的 Yohkoh 数据库。随后又有很多单位加入了提供数据服务的此行列。表 1 列出了 VSO 的数据组成情况。表 1 中: 1) KPVT 即 Kitt Peak Vacuum Telescope; 2) OSO-7 即 The Seventh Orbiting Solar Observatory; 3) SMM 即 The Solar Maximum Mission; 4) CGRO BATSE 即 Compton Gamma-Ray Observatory Burst and Transient Source Experiment; 5) SOHO 即 Solar and Heliospheric Observatory; 6) TRACE 即 Transition Region and Coronal Explorer; 7) GSFC/SDAC 即 Goddard Space Flight Center/Solar Data and Analysis Center; 8) SOI 即 Saturn Orbit Insertion; 9) SSSC 即 the Stanford SOI Science Support Center; 10) SXT 即 Soft X-ray Telescope; 11) USC 即 University of Southern California; 12) BBSO/NJIT 即 Big Bear

Solar Observatory/the New Jersey Institute for Technology ; 13) PSPT 即 Precision Photometric Solar Telescope ; 14) ECHO 即 Experiment for Coordinated Helioseismic Observations ; 15) ASP 即 Advanced Stokes Polarimeter ; 16) CHIP 即 Chromospheric Helium I Imaging Photometer ; 17) HAO 即 The High Altitude Observatory 。

表 1 VSO 的数据组成^[11]

数据档案	归 属
KPVT ¹⁾ , GONG, SOLIS	The NSO Digital Library
OSO-7 ²⁾ , SMM ³⁾ , Yohkoh, CGRO BATSE ⁴⁾ , SOHO ⁵⁾ , TRACE ⁶⁾	NASA/GSFC SDAC ⁷⁾
SOI/MDI ⁸⁾ , TON, Wilcox Solar Observatory	The Stanford University SSSC ⁹⁾
MDI, TRACE, Yohkoh/SXT ¹⁰⁾ , LaPalma archive	Lockheed
TRACE, Yohkoh/SXT, the Mees Solar Observatory	Montana State University
Helioseismology Data Set of Dopplergrams	USC/Mt. Wilson 60-ft Fower ¹¹⁾
Full-disk Halpha, Ca K and White Light Images	BBSO/NJIT ¹²⁾
ARTHEMIS	Italy
BASS2000	France
Coronal, PSPT ¹³⁾ , ECHO ¹⁴⁾ , ASP ¹⁵⁾ , CHIP ¹⁶⁾	HAO ¹⁷⁾

1.4 设计、技术概念及其发展

(1) 设计

VSO 的设计目的是使太阳物理学家能用浏览器接口在 VSO 中搜索他们用于解决研究问题的可用数据, 通过精化搜索标准, 直接从数据源检索数据, 避免通过中间环节搜索可能遇到的网络堵塞。

VSO 允许用户使用任何支持网络连接的语言 (如 C、C++、Java、Perl、Python、IDL SolarSoft 树、IRAF) 来开发自己的接口及处理程序, 使用户能将自己的开发连入 VSO, 进行本地数据分析。原型阶段的设计将提供: 数据搜索、数据发现、查询精细化; 多种用户接口 (可浏览的、应用程序接口)。

以美国 VSO 设计为例, 其图形用户查询界面初步由以下几部分组成: 数据源、观测内容、维数、波长、时间、空间等; 各种可替换的数据格式和数据传输方法。其中数据源包括表 1 中 10 个主要的太阳数据源; 观测内容包括: 强度、多普勒速度场、矢量磁场、纵向磁场、斯多克斯参数 (I 、 Q 、 U 、 V)、粒子流、射电流、声功率、振荡参数、太阳黑子数; 维数由 X (东西方向)、 Y (北南方向)、 Z (高度)、 T (时间)、波长、球谐度、方位角度、径向等级、时间频率选项组成; 查询返回结果则包含: 某时间不同地点、不同谱线的多个观测图像。数据格式有 XML (eXtensible Markup Language)、HDF (Hierarchical Data Format) 或 FITS (Flexible Transport System)。所有在开发中的特性, 如太阳特性目录、太阳事件目录等都将成为 VSO 的服务内容。

(2) 技术概念

技术概念包括: 分布太阳档案联合、适应性元数据词典、单一化且直观的图形用户查询界面、基于内容的检索、分布式计算。潜在的工具包括: 分析软件、数据可视化、可视化编程接口及使用追踪系统。原型阶段运作完成后, 数据目录缓存功能、实例、搜索记录 (追踪系统) 再

被加入。其中元数据词典是建立全世界统一的太阳数据模型,避免使用者对数据进行本地化处理;分析和带有搜索能力的图形软件包的结合将产生新的强有力的太阳研究工具;分布式计算需考虑在查询时分布在可用服务器之间的动态负载平衡;实例是很多用户渴望的特性,它对于传输记录信息到主服务器是必需的,以便搜索记录被记录下来;追踪系统用以评价 VSO 对太阳物理的全面影响,如效能和产出、搜索结果再生等,它对如何废弃数据是一个很好的评价实践,同时也支持对基金投入的评价,使对同样数据使用不同分析方法成为可能。

这个系统底部结构使用与硬件平台无关的工具如 XML 和 JavaScript,来完成数据服务描述和服务合成。分布计算使用网格技术。在 VSO 的开发中将面临一些技术挑战,如它的安全发布、网络带宽、元数据表示、网格技术等。由于不断有数据集和分析软件包加入到 VSO 中,VSO 是可扩展的系统。

(3) 技术概念的发展

VSO 正不断从数据提供方得到程序支持,以实现专门数据对 VSO 的接入和整个 VSO 系统的实现。对应于 VSO,还要考虑^[11]:在存储和处理数据或本地没能力通过 Internet 传输大量数据时,可通过使用智能代理在分布档案所在地去识别和处理大量数据采样,这样即使是没有经验的用户,只要简单地接受他人的分析过程也可以完成检索并获取数据。

VSO 性能扩展依赖于研究需求,如可将 EGSO (European Grid of Solar Observation) 的特性和事件列表联入 VSO 中;增加远程处理能力,减少数据传输量(通过基于数据的特性或事件的存在与否来自动选择);以即插即用方式联到 LWS (Living With a Star,这是致力于科学理解并有效解决直接影响人类生命和社会的与日地系统相关问题的项目)系统中。

2 虚拟太阳天文台的发展

VSO 概念最早由美国提出,随后得到了全世界太阳物理界的积极响应和运作。

2.1 中 国

中国是发展中国家,其天文观测设备尤其是空间观测设备的发展资金不足,研制技术还需一个相当长的储备过程,因此,良好的数据获取途径及其相应的进一步的数据分析处理工具显得尤为重要。为了适应研究需要,使研究工作走在世界前沿,弥补空间太阳观测资料上的空白,早在上世纪 90 年代初,我国太阳物理学界就与国外相应机构开展了联合观测,并交换地基与空间资料。可见,发展我国 VSO 意义重大。

中国科学院国家天文台怀柔太阳观测基地,以观测太阳磁场、速度场和太阳爆发活动等太阳物理领域中的前沿研究课题为任务,其良好的矢量磁场资料受到国际同行的高度重视,并促成了中国太阳物理界与美国、日本、俄罗斯、法国、德国等国签署正式的太阳物理合作研究计划。多年来,怀柔太阳观测基地成功地与国际上著名观测站 SOHO、TRACE、Yohkoh 等进行了联测;与日本进行了地基观测资料与空间资料的交换;与美国大熊湖天文台开展了“日不落”联合观测;参与了怀柔、昆明、大熊湖、奥地利 H-alpha 全日面图像的全球观测网,与大熊湖一起在各自网站上发布每天的观测资料;形成了以太阳磁场观测为主体的良好的国际联系和研究氛围。已建成的 WWW 数据网站包括磁场、全日面 H-alpha、元数据、数据格式说明、数据处理程序等内容。它的数据形式为国际标准的 FITS 格式及处理过的图像,另外

还提供对最新数据最快捷的浏览方式, 这为建立我国的 VSO 打下了基础。其中部分原数据和处理过的数据已提供在线服务, 用户可通过 http 或不具名 FTP 得到 (但检索功能未提供), 更多的数据存储于 MO (Magneto-Optical) 及光盘上。相信, 在硬软件上的不断更新改造以及新研制的观测仪器的使用将会使其数据种类增加, 质量变得更为优异。

2002 年初, 在国际天文界提出建立“全球虚拟天文台”(Global Virtual Observatory, GVO) 时, 以中国科学院国家天文台为首的中国天文界与高等院校的天文机构沟通, 不甘落后^[12], 也提出了建设“中国虚拟天文台”的设想, 着手创建中国虚拟天文台, 并成为国际虚拟天文台联盟 (International Virtual Observatory Alliance, IVOA) 中的一员, 目前已初步完成了体系结构设计工作。与此同时, 中国 VSO 也有了初步设计, 有关情况详见文献 [13], 这里不再赘述。

2.2 其他国家或组织

意大利提出了 SOLARNET (SOLAR NETWORK) 项目^[14], 目的是集中所有意大利太阳档案, 形成 IVSO (Italian Virtual Solar Observatory) 并作为 EGSO 的一部分。意大利国内分布的太阳档案有: 位于 Torino 的由 INAF (Istituto Nazionale Di Astrofisica) 天文台维护的 SOHO Long-term Archive; 位于 Napoli 的由 INAF 天文台维护的 THEMIS Archive、罗马 INAF 天文台数据及 Catania INAF 天体物理天文台数据; 位于 Triest 的太阳射电档案, 此档案也是通向奥地利 Kanzelhoehe Solar Observatory 的数据通道。SOLARNET 允许查询所有上述连接的数据档案, 它是更为大型和高级项目 EGSO 的试验床, 为实现世界范围的 VSO 打下了基础。

欧盟提出了 EGSO 项目^[15]。这是基于网格技术, 用于解决分布地异构数据集、显著地改善分散用户对太阳数据访问的项目。它的主要任务是, 面对元数据在质量、内容、相关性和格式上的不同, 产生一套新的统一观测目录、太阳特性目录和太阳事件目录。其中太阳特性目录由一系列发生、变化的太阳特性数据组成, 这些数据包括: 暗条、黑子、活动区、日冕物质爆发等; 太阳事件目录由公布的事件列表组成, 包括: 人工生成目录 (如 SOHO CME, 即 Coronal Mass Ejection 列表)、来自 SGD (Solar-Geophysical Data) 的耀斑目录等。后两个目录将提供基于太阳特性、事件、现象的新切入点 (不只是日期、时间、波长和指向)。在数据处理上包括数据抽取和定标, 使用 SolarSoft 及专业软件进行数据比较和分析。在分析太阳数据中, 主要障碍是找出可用数据和检索必需数据。同时还要考虑空间和地面仪器产生的巨大数据量以及技术的不断进步。EGSO 将改变用户分析太阳数据的方式, 把欧洲内外的太阳数据档案联合起来, 生成能挑选、处理、检索分布和异构太阳数据的工具, 另外还将提供产生空间和地基的标准观测目录机制, 这一机制能使挑选基于特性、事件和现象的太阳数据变得容易。实际上 EGSO 构造了一个虚拟天文台。

美国国家虚拟天文台 NVO 比 VSO 更有名, 它将大范围的天文数据库、数据挖掘和其他关注的特性提供给夜间工作的天文学家。

日本、印度、韩国等众多国家也都提出了各自的 NVO 或 VSO 计划。

3 VSO 的作用与意义

VSO 的使用将促进科学研究的进一步发展, 有些科学问题在 VSO 出现以前是很难开展研究的。譬如:

(1) 空间天气学

通过相关时间序列日冕图像、暗条位置、矢量磁场、X 射线等多波段观测资料, 预报由耀斑和日冕现象(如日冕物质抛射)引起的地磁暴。

(2) 太阳黑子大气精细结构

由太阳矢量磁场图、空间观测的紫外光谱、极紫外图像的时间序列组成的数据集, 能对太阳黑子大气密度、磁场、速度场、扰动等作更完整的描述。

(3) 活动区亚表面结构

该结构最早是通过日震学来探究的, 日震学通过太阳内部的声震动推断太阳内层特性。这种研究需要综合表面磁场测量和长时间序列多普勒速度场测量数据。

(4) 太阳风

通常认为, 太阳风产生于冕洞周围。在冕洞附近, 太阳磁场对行星际空间开放, 太阳风从太阳上传递带电粒子到地磁场和大气中。目前太阳风物理加速过程还不清楚, 这需要利用能反映太阳风速和风量的 X 射线或红外波长数据得到的图像来进行研究。

还有很多需要用多波段观测进行科学研究的例子, 这里不再一一列举了。作为查询工具, VSO 开拓了跨多维和多个数据源的大尺度的相关统计太阳物理新领域, 它将直接应对太阳研究的主要需求之一: 定位、关联、获得和分析来自超过 7 个量级的时空尺度的望远镜观测数据。

日地系统研究迫切需要 VSO。VSO 将极大地增加日地系统相关研究的能力和加快相关研究的步伐。负有太阳物理使命的空间望远镜, 如 Yohkoh、SOHO 和 TRACE 已经获得了大量数据^[16], 其中 SOHO、TRACE 还将继续获取数据。在下一个 5 年中, 随着 Solar-B、STEREO (Solar Terrestrial Relations Observatory)、SDO (Solar Dynamics Observatory) 空间望远镜的投入使用, 观测数据量将剧增。目前很多数据库都有镜像, 即这些数据库分布在 Internet 上的多个地点。类似地, 基于地面的日震网络如 GONG+, 正在以比以往任何时候都快的速率传递数据。SOLIS 综合观测计划承诺增加比 Pb (1 Pb = 10⁶ Gb) 量级还多的数据; 2007 年 SDO 每天将产生 1 Tb 元数据。数据源存储量及多样性的增加, 会在技术上带来不可忽视的费用。定位所有服务器和理解不同查询工具也将变得更加困难。目前, 原有的极端耗时的劳动密集型数据查找方式妨碍和影响了很多研究人员对于太阳结构、太阳活动机制和变化进行多波段综合研究的进程。

简化不同服务节点的多样性查询是 VSO 的功能之一。VSO 是 Internet 上不断扩大、数据整合、可进行数据分析的环境。通过数据最大程度地被使用, VSO 将会极大地使可观的投资设备升值。在技术方面, 存储技术的发展使原先劳动密集型的存储状况消失, 存储设备既能满足 Pb 量级数据的存储(所需费用并不太多), 也能将集中化的数据服务转变为分布式服务, 后者更具有使长期数据免遭破坏的优势。对数据服务方式发展的洞察力使太阳物理界达成了共识: 一个统一的基于网络分布式访问的系统可以替换集中式的数据中心, 这将更适合太阳物理学研究。VSO 技术建立在分布式数据访问和分布式计算的基础上, 它的建立和实现将为推动网格技术的发展、应用作出贡献。

此外, VSO 还将为教育和科普作出贡献。由于太阳对人类社会生活具有支配作用^[11], 太阳在科学教育事业上长期扮演着重要角色, 天文、物理、生命科学、艺术、甚至音乐(通过日震学)等学科都可以从中受益。VSO 将拥有完整的太阳科学教育课程模块, 为非专业用户提供第二个数据通道, 这个“教育通道”是图像型的, 主要通过图像方式选择数据。

4 结 语

VSO 的执行与实施将建立全球统一的太阳数据, 成为对全球太阳数据统一查询、获取、比较和处理的理想工具, 并产生专门的太阳特性目录和太阳事件目录。每个虚拟天文台都具有各自的天文应用和技术特色, 针对专门的太阳物理研究领域 (天文学研究), 提供前所未有的手段和方法。VSO 将有力推动新的研究方法和新的多波段探测方法的发展, 全球太阳数据将被统一, 未来的空间望远镜 (如 STEREO、Solar-B、SDO) 的数据都将通过 VSO 提供给用户使用, 而不需要先前意义上的中心服务器。从此虚拟太阳天文台的时代就开始了。

参考文献:

- [1] Rouder Th, Malherbe J M. *Obs. de Paris Publ.*, 1997, 45: 6
- [2] Leka K D. *Sol. Phys.*, 1994, 155: 301
- [3] Sanchez-Duarte L, Fleck B, Bentley R. *ASP Conf. Ser.*, 1997, 118: 382
- [4] <http://arthemis.na.astro.it/wsc/proposal/cnaa.proposal.html>, 2002
- [5] Dimitoglou G, Mendiboure C, Reardon K *et al.* *ASP Conf. Ser.*, 1998, 155: 297
- [6] <http://vso.nso.edu/vsowp.pdf>, 2000
- [7] Gurman J B. In: Wilson A ed. *Proc. SOHO 11 Symp.*, SP-508, Beitrag: ESA, 2002: 66
- [8] Messerotti M, Veronig A. *ASP Conf. Ser.*, 1998, 155: 462
- [9] Cora A, Volpicelli C A, Antonucci E *et al.* *Mem. S. A. It.*, 2003, 74: 462
- [10] Mendiboure C. *ASP Conf. Ser.*, 1998, 155: 302
- [11] Hill F. *ASP Conf. Ser.*, 1998, 155: 297
- [12] 张彦霞, 赵永恒. *天文学进展*, 2004, 22(4): 350
- [13] 林钢华. *国家天文台台刊*, 2004, 1(3): 210
- [14] Messerotti M, Coretti I, Padpvan S *et al.* *Mem. S. A. It.*, 2003, 74: 391
- [15] Bentley R D. In: Alissandrakis C E, Schmieder B eds. *Proc. 10th. European Solar Physics Meeting*, SP-506, Beitrag: ESA, 2002: 26
- [16] http://Virtualsolar.org/docs/VSO_strawman_20021125.pdf, 2002

The Virtual Solar Observatory and its Development

LIN Gang-hua

(National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China)

Abstract: With going deep into research of solar physics, development of observational instruments and accumulation of observational data, it urges people to think such things: using data which is observed in different times, places and bands to seek answers of a plenty of scientific problems which are hang in doubt. In the meanwhile, researcher can easily search, analyze and use data. That is why the project of the virtual solar observatory (VSO) gained active replies and operations from observatories, institutes and universities in the world. In this article, the origins, function, adopted technique and the current status on the virtual solar observatory are introduced.

Key words: astronomy; VSO; review; solar physics; data archive; search tool online; network; grid