

OBSERVATORIOS VIRTUALES ASTROFÍSICOS

M. en I. Liliana Hernández Cervantes

Jefa de la Unidad de Astrofísica Computacional del Instituto de Astronomía de la UNAM

liliana@astroscu.unam.mx

Dr. Alfredo Santillán González

Investigador Titular de la Dirección General de Servicios de Cómputo Académico de la UNAM

alfredo@astroscu.unam.mx

M. en I. Alejandro R. González-Ponce

Jefe de la Unidad de Cómputo del Instituto de Ecología de la UNAM

alex@ecologia.unam.mx

Observatorios Virtuales Astrofísicos

Resumen

Uno de los principales retos que tiene actualmente la investigación en astrofísica es el manejar y analizar el crecimiento exponencial y la complejidad de datos observacionales, junto con el enorme volumen de resultados computacionales, que se han generado en todo el mundo en las últimas décadas. El concepto de Observatorio Virtual (OV) es la mejor respuesta a este problema, ya que permite el acceso transparente y distribuido de los datos a través de Internet, además de proporcionar poderosas herramientas para analizarlos y visualizarlos. Aquí presentamos una descripción general del concepto de OV, y damos algunos ejemplos de trabajos hechos en varios países, incluyendo México.

PALABRAS CLAVE: Observatorios Virtuales, Bases de Datos, Minería de Datos, Simulaciones Numéricas, Astrofísica

Abstract

One of the main challenges in today's astrophysical research is handling and analyzing the exponential growth and complexity of observational data, along with the large volume of computational results, that has been generated worldwide over the last decades. The Virtual Observatory (VO) concept is the best answer to this problem that allows for a transparent and distributed access to the data via internet, and also provides powerful tools to analyze and visualize them. Here we present a general overview of the VO concept, and give some examples of the work done in several countries including Mexico.

KEYWORDS: Virtual Observatories, Data Base, Data mining, Numerical Simulation, Astrophysics

Introducción

Al igual que en otras ciencias y en la mayoría de las actividades que están vinculadas al desarrollo moderno de nuestra sociedad (por ejemplo, comercio, entretenimiento, industria, seguridad, etc.) la astronomía está siendo abarrotada de información por un crecimiento exponencial en el volumen y la complejidad de datos observacionales y teóricos que se van generando en todo el mundo (Szalay & Gray 2000). Se estima que la cantidad de información acumulada se duplica cada 18 meses y por lo tanto el crecimiento en volumen cumple perfectamente con la ley de Moore (1965). Esto no es una casualidad, ya que mientras el avance del área de los espejos primarios de los telescopios, a lo largo de los años se ha duplicado aproximadamente cada 25, el número de píxeles de un detector CCD (*Charged Coupled Device*) se duplica cada 2 años (ver figura 1). Esto quiere decir que aún cuando el tamaño de los telescopios permanezca fijo durante cierto tiempo, el tener detectores CCDs más grandes y más sensibles, implica generar archivos cada vez de mayor tamaño. Esto trae como consecuencia un serio desafío a la comunidad científica internacional, ya que la cantidad de información astronómica a la que se enfrenta los grupos de investigación es realmente muy grande. Por ejemplo, hace apenas 7 años ya se calculaban el contenido de bases de datos astronómicas en varios cientos de TB (1 *Terabyte* = mil *Gigabytes* = un millón de *Megabytes*) (Brunner *et al.* 2002) y a principios de 2005 la tasa de recolección de datos se estimaba aproximadamente a un *Terabyte* (TB) por día. Entonces, dado que tanto el volumen de datos así como sus tasas de recolección están creciendo exponencialmente, nos lleva a pensar que en pocos años nos estaremos enfrentando a bases de datos de varios PB.

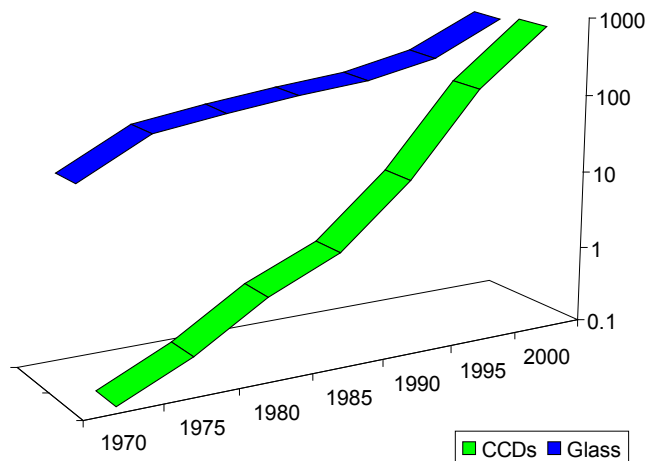


Figura 1. La gráfica muestra el crecimiento de los espejos primarios de los telescopios y de las dimensiones de los detectores CCD respecto al tiempo. Claramente se puede ver que el aumento de la resolución de los CCDs cumple con la ley de Moore, es decir, presenta un crecimiento exponencial.

Telescopios: generadores de grandes bases de datos.

Hasta la fecha existen telescopios terrestres y espaciales produciendo decenas de TB de datos astronómicos, como por ejemplo, el *Hubble Space Telescope (HST)*, *Chandra X-Ray Observatory*, *Sloan Digital Sky Survey (SDSS)*, *Two Micron All Sky Survey (2MASS)* y *Faint Images of Radio Sky at Twenty Centimeters (FIRST)*, por mencionar algunos. Todos estos instrumentos, junto con los nuevos telescopios en construcción como el *Large Synoptic Survey Telescope (LSST, <http://www.lsst.org/lsst>)*, seguirán produciendo una enorme cantidad de información que estará almacenada en diferentes bases de datos del orden de *Petabytes (PB)*. La cantidad de información astronómica a la que nos estamos enfrentando es realmente descomunal. Por ejemplo, en mayo de 2009, sólo el proyecto SDSS tenía ~60 TB de información de millones de objetos astronómicos (asteroides, estrellas, cúmulos de estrellas, nebulosas, supernovas, galaxias, cúmulos de galaxias, etc.) observados en diferentes longitudes de onda. En tanto que en el futuro el proyecto LSST producirá del orden de varios cientos de MB por segundo, es decir, más de 8 TB por día, cantidad de información que será difícil de dirigir sin herramientas computacionales de vanguardia. Por otro lado, el desarrollo futuro de la astrofísica mexicana y de las ciencias computacionales en general dependerá de las posibilidades de crear nuevos proyectos de gran envergadura, que permitan construir una infraestructura sólida y que estimulen el crecimiento de las plantas científicas y técnicas. Recordemos que de forma conjunta, la Universidad Nacional Autónoma de México (UNAM) y el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), cuentan con un par de Observatorios en Baja California y Sonora, y están construyendo el Gran Telescopio Milimétrico (GTM) y el observatorio de rayos gamma HAWC en Puebla. De igual manera ambas instituciones son miembros del Gran Telescopio de Canarias (GTC), de los radiointerferómetros *Expanded Very Large Array (EVLA)* y *Atacama Large Millimeter Array (ALMA)* y parte del grupo promotor del satélite *World Space Observatory*.

¿Qué es un Observatorio Virtual Astrofísico?

En la sección anterior se dio una descripción general de la cantidad de información que se está generando, particularmente; en astronomía, pero que también se produce en otras áreas de la ciencia (biología, ciencias genómicas, física de altas energías, química, etcétera.). Queda claro que una vez que se tengan almacenados estos volúmenes de datos, será imposible acceder, manipular y analizar de manera eficiente toda la información contenida allí, a menos que contemos con la infraestructura y la planeación adecuada. El origen del concepto de **Observatorio Virtual (OV)** surge de una propuesta científica-tecnológica que tiene la

intención de dar respuesta al enorme problema de operar adecuadamente la información masiva, producida por diferentes medios que se utilizan para estudiar el Universo, telescopios y modelos numéricos. Un **OV** ofrece un *ambiente federativo internacional virtual de investigación*, basado en nuevas tecnologías de la información e Internet, completamente abierto a científicos y estudiantes que desean trabajar con conjuntos de datos astronómicos reales. Esta herramienta computacional reúne archivos de datos y servicios, así como complejas técnicas de exploración (minería de datos) y análisis de datos. Un **OV** es un excelente pretexto para realizar proyectos multidisciplinarios donde colaboren astrofísicos y especialistas en ciencias e ingeniería de la computación (Djorgovski & Williams, 2005). Y por si fuera poco, dicho concepto puede extenderse y aplicarse fácilmente a otras áreas de la ciencia y de la sociedad en general, que generan y almacenan datos de manera frecuente, como es el caso del Instituto Nacional de Estadística, Geografía e Informática (INEGI).

La definición de **OV** fue propuesta en la década de los años noventa a través de un sinnúmero de discusiones y talleres realizados durante el Simposio 179 de la Unión Astronómica Internacional y en la Reunión 192 de la *American Astronomical Society* (Djorgovski, & Beichman, 1998). En dichas discusiones se acordó que el acceso electrónico, vía Internet, a las bases de datos astronómicos de todo el mundo, es vital para investigar los detalles del Universo. Una base de datos por si sola es importante, no obstante, si se desea realizar investigaciones que requieran un estudio multi-espectral, o pancromático, se debe tener acceso a varias bases de datos.

Así pues, los observatorios virtuales tienen como una de sus tareas fundamentales dar acceso a los astrónomos mexicanos y de todo el mundo a la información almacenada en dichas bases de datos, independientemente del lugar geográfico donde se generen o almacenen los datos. Por ejemplo, si se quiere entender los detalles de la expansión del Universo y la formación de la galaxias, se necesitan estudios estadísticos de tipos específicos de galaxias, así como de sus distancias y el medio ambiente donde se localizan. Esto requiere de imágenes en diferentes longitudes de onda, de miles o millones de galaxias, así como el conocimiento de sus distancias. Esta tarea resultaría casi imposible de lograr con una sola base de datos.

Arquitectura de los Observatorios Virtuales

Como ya se mencionó, un **OV** es una herramienta computacional que permite almacenar, acceder y procesar información de manera eficiente en grandes bases de datos, distribuidas alrededor de todo el mundo, utilizando intensivamente la Internet para operar sobre los datos. De manera general se puede decir que normalmente están compuestos por varios módulos, como los asociados a: la adquisición y la operabilidad de los datos, la generación de las bases de datos, las herramientas de búsqueda en las distintas bases, el diseño y el desarrollo de las herramientas de reducción y procesamiento, los relacionados con la visualización, etcétera.

El éxito de los **OV's** radica en la capacidad de interactuar entre las distintas bases de datos, de manera sencilla desde una computadora de escritorio conectada a la red. La Alianza Internacional de Observatorios Virtuales (IVOA por sus siglas en inglés) es la entidad federativa responsable de dictar los estándares sobre la forma de "operar" los **OV's**. Estos estándares van desde la manera en la que se "etiquetan" los datos (generar el metadato), hasta la implementación de herramientas computacionales que se aplicarán sobre ellos. Existen diferentes grupos de trabajo en la IVOA que se encargan de generar estos estándares, los cuales están relacionados con las aplicaciones, los sistemas de búsquedas, la forma de comunicar las bases de datos, la modelación de los datos y la generación de reglas para los metadatos, entre otros. La IVOA agrupa a 15 países (Alemania, Armenia, Australia, Canadá, China, España, Estados Unidos, Francia, Hungría, India, Italia, Japón, Korea, Reino Unido y Rusia,) y a la Comunidad Europea (ver figura 2). Algunos de los miembros cuentan con telescopios terrestres o naves espaciales, como es el caso de Estados Unidos, pero otros no, tal como la India; sin embargo, todos trabajan en el desarrollo de diversas herramientas computacionales para operar las bases de datos, que están disponibles para la comunidad científica.



Figura 2. Alianza Internacional de Observatorios Virtuales

Afortunadamente, ya es posible contar con un gran número de herramientas que facilitan la manera de llevar a cabo la minería de datos, la cual consiste, literalmente, en “extraer” la información valiosa de una “montaña de datos”. Existen aplicaciones desarrolladas por el grupo de programadores de la IVOA, como TOPCAT, por mencionar alguna, que mediante una interfase gráfica interactiva proporciona muchas de las facilidades que los astrónomos necesitan para analizar y manipular datos (ver figura 3); por ejemplo, maneja diferentes formatos de datos como FITS y VOTable, además de ofrecer diferentes formas de visualizarlos y analizarlos. Cabe mencionar que una de sus principales fortalezas es la de acceder rápidamente a grandes volúmenes de datos.

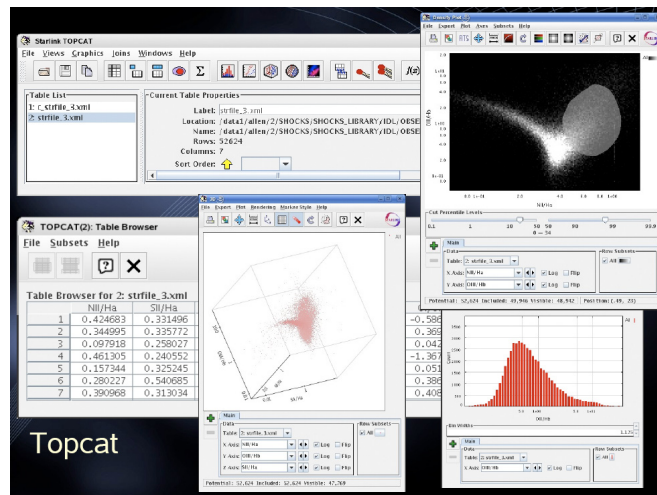


Figura 3. Ejemplo del tipo de interfases que utiliza Topcat (Mark Allen)

Observatorios Virtuales Teóricos

Es bien sabido que las simulaciones numéricas juegan un papel fundamental para estudiar y entender la evolución de diferentes eventos astronómicos que ocurren en el Universo, ya que estos fenómenos cósmicos necesitan de miles o millones de años para desarrollarse y sería imposible llevar un seguimiento puntual

debido a que los seres humanos viven en promedio 80 años. Hoy en día, gracias al uso de equipos de cómputo de alto rendimiento y al desarrollo de complejos códigos numéricos astrofísicos, es posible llevar a cabo simulaciones numéricas astrofísicas con gran detalle a muy alta resolución. Esto trae como consecuencia que la cantidad de datos generados por cada una de estas simulaciones, sea muy grande, no sólo por la simulación numérica resultante, sino porque para llegar al resultado deseado, en ocasiones se requiere un sinnúmero de cálculos numéricos previos.

El concepto de Observatorio Virtual Teórico (OVT), está estrechamente relacionado con un OV, sólo que en lugar de almacenar datos obtenidos con telescopios, está conformado por datos generados por modelos astrofísicos teóricos. El objetivo de este tipo de observatorios es proporcionar a la comunidad científica, con poca o nula experiencia en códigos numéricos, una serie de herramientas computacionales que le permitan realizar simulaciones numéricas, ejecutando remotamente un código, o bien utilizando datos de cálculos numéricos que ya han sido ejecutados con anterioridad, aprovechando las características que ofrece el envío de datos por *Internet* y en su caso las redes de alta velocidad.

En el 2004 se creó la “*IVOA Theory Interest Group*”, el cual tiene como objetivo principal garantizar que los datos teóricos estén considerados en los procesos de estandarización del Observatorio Virtual. Una de sus principales tareas es desarrollar herramientas que permitan comparar los resultados teóricos con las observaciones y viceversa. Algunos ejemplos de OVT que se han realizado en otros países, se pueden ver en la tabla 1.

Observatorio	País
<u>Cosmo.Lab</u>	Italia
<u>Modelo de Síntesis de Población Estelar de nuestra Galaxia.</u>	Francia
<u>Proyecto Virgo</u>	Alemania

Tabla 1. Ejemplos de algunos Observatorios Virtuales Teóricos

Observatorio Virtual Solar Mexicano

En el caso particular de México, la UNAM, a través de la Dirección General de Servicios de Cómputo Académico y los Institutos de Astronomía y Ecología, desarrolló el primer Observatorio Virtual Solar de nuestro país, el Observatorio Virtual Solar Mexicano (OVSM). Este fue diseñado a partir de la premisa de crear una herramienta computacional sencilla de utilizar, que cumpliera con los requisitos básicos de un Observatorio Virtual, donde la base de datos se genere automáticamente con resultados de simulaciones numéricas, ejecutadas remotamente en un servidor, y no con datos observacionales. Sus características hacen que sea una excelente opción para que estudiantes o investigadores que trabajan con datos observacionales, cuenten con un modelo hidrodinámico opcional que les permita interpretar sus datos. Es importante mencionar que esta herramienta computacional está diseñada para académicos que no cuentan con equipos de cómputo de alto rendimiento o no son expertos en códigos numéricos. Todo se hace a través de una interfase *Web* intuitiva. Hasta el momento el OVSM está orientado al estudio de la Evolución de Eyecciones de Masa Coronal (EMC) en el Medio Interplanetario, tema de gran interés para investigadores dedicados a estudiar problemas vinculados al **clima espacial**. Éste es un conjunto de fenómenos e interacciones que se desarrollan en el medio interplanetario, que está regulado fundamentalmente por la actividad que se origina en el Sol, y nos proporciona información sobre el estado de perturbación del ambiente entre la Tierra y el Sol (ver figura 4).

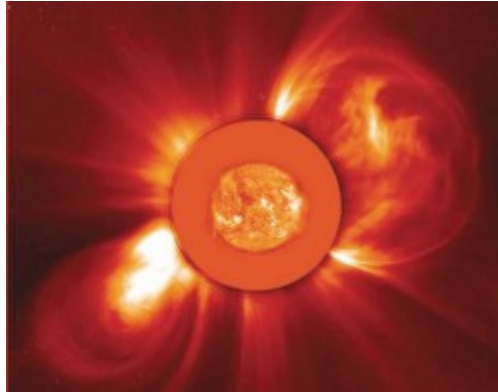


Figura 4. En esta imagen se muestran 2 EMC simétricamente opuestas. Las observaciones se realizaron el 8 de Diciembre de 2000 con la nave espacial SOHO, utilizando el coronógrafo LASCO-C2. Una imagen del EIT, tomada el mismo día, ha sido sobrepuesta en lo que sería el coronógrafo C2, el cual es usado para ocultar la luz fotosférica y poner así de manifiesto la débil señal coronal en luz blanca. (*Solar & Heliospheric Observatory, SOHO*).

La arquitectura computacional del OVSM está basada en tres módulos. El primero corresponde al diseño de la Interfaz gráfica de usuario (GUI por sus siglas en inglés); el segundo está relacionado con la ejecución de la simulación numérica remota, y el tercero está vinculado con la generación de la base de datos. Cada uno de estos módulos está vinculado entre sí (ver Hernández-Cervantes, *et al.* 2008). Un ejemplo del tipo de resultados que se pueden obtener con el OVSM.

Algunos ejemplos de otros Observatorios Virtuales Solares que se han desarrollado, se muestran en la tabla 2.

Observatorio	País
<u>NVSO</u>	Estados Unidos
<u>VESO</u>	México
<u>VSTO</u>	Estados Unidos

Tabla 2. Ejemplos de otros Observatorios Virtuales Solares

Conclusiones

El desarrollo de la astronomía está íntimamente ligado a los avances tecnológicos. En el caso de los Observatorios Virtuales, las nuevas tecnologías de la información han facilitado a la comunidad científica el manejo de los enormes volúmenes de datos. Éstos han sido generados por telescopios y han sido arrojados en diferentes simulaciones numéricas, proporcionándoles una excelente herramienta para hacer investigación de frontera, además de ofrecerles la posibilidad de combinar la observación con la teoría y viceversa, con el

objeto de entender el desarrollo de nuestro Universo.

Agradecimientos:

Este trabajo fue parcialmente financiado por el proyecto PAPIIT IN121609-3 de la DGAPA, UNAM.

Bibliografía

Brunner, R., Djorgovski, S.G., Prince, T., & Szalay, A. 2002, in: *Handbook of Massive Data Sets*, eds. J. Abello et al., Dordrecht: Kluwer Academic Publ., p. 931.

Chandra X-Ray Observatory, <http://physics.nist.gov/PhysRefData/Chandra/index.html>

Djorgovski, S.G., & Beichman, C. 1998, *BAAS*, 30, 912.

Djorgovski, S. G.; Williams, R., 2005, *From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science ASP Conference Series*, Vol. 345, edited by N. Kassim, M. Perez, M. Junor, and P. Henning, p.517-530

Faint Images of Radio Sky at Twenty Centimeters, <http://sundog.stsci.edu/>

Hernández-Cervantes, L., González-Ponce A., Santillán A. and Salas G., *Computational backbone of the Mexican Virtual Solar Observatory*, VII Conferencia Latinoamericana de Geofísica Espacial, Mérida, Mérida, Yucatán, 2008.

Hubble Space Telescope, <http://www.stsci.edu/hst/>

International Virtual Observatory Alliance, <http://www.ivoa.net/>

Large Synoptic Survey Telescope, <http://www.lsst.org/lsst>

Mark Allen, *Overview of VO applications*, "Astronomy with Virtual Observatories", Pune, India, Oct 15, 2007

Moore, G., 1965, *Cramming More Components onto Integrated Circuits*, *Electronics Magazine* Vol. 38, No. 8.

Observatorio Virtual Solar Mexicano, <http://mvso.astroscu.unam.mx>

Sloan Digital Sky Survey, <http://www.sdss.org/dr3/index.html>

TOPCAT, <http://www.star.bris.ac.uk/~mbt/topcat/>

Two Micron All Sky Survey, <http://www.ipac.caltech.edu/2mass/>