Astrophysical Integrated Research Environment

Jianfeng Zhou[1], Rui Qi[2], Junwei Si[1] & Jianying Nie[1]

(1 *Center for Astrophysics, Tsinghua University, Beijing* 100084)
(2 *Department of Automation, Tsinghua University, Beijing* 100084)

Abstract One big challenge that modern astrophysicists face is how to use Terabytes (maybe Perabytes soon) of astrophysical data effectively. The proposed AIRE (Astrophysical Integrated Research Environment) plan is to help astrophysicists solve the problem. The AIRE consists of three main parts: Data Archive Center (DAC) which collects and manages public astrophysical data; a web-based Data Processing Center (DPC) which enables astrophysicists to process the data in a central server at any place and anytime; and a Concurrent Research System (CRS) with which astrophysicists in different fields can pursue a collaborative reserch efficiently.

1 Introduction

Since the application of digital apparatuses like CCD in telescopes, astrophysical data have increased dramatically. At present, the total amount of data is about several hundred Terabytes. They cover from radio band to γ-ray band. In near future, several next generation telescopes, such as LASST (Large-Aperture Synoptic Survey Telescope), will be lunched. By then, over 10 Terabytes of astrophysical data will be generated each day. So, the total amount of data will exceed easily Perabyte ($1P = 10^{15}$).

Such huge data will probably lead to new discoveries in astrophysics. However, how to mine the knowledge hidden in these data is a big challenge for astrophysicists. Consequently, virtual observatory concept has been proposed (Boroson et al. 2001, Brunner et al. 2001). At present, there are several VO projects around the world, including NVO (National Virtual Observatory) in United States, AVO (Astrophysical Virtual Observatory) in Europe, and Austrilian Virtual Observatory.

In China, we have no VO project yet. However, considering that China is building LAMOST (Large Sky Area Multi-Object Fiber Spectroscopic Telescope) and will presumably launch HXMT (Hard X-ray Modulation Telescope) and other space and ground telecopes in future, it is worthwhile to develope some VO techniques in advance. In order to facilitate the research work of Chinese astrophysicists, it is urgent to collect public astrophysical data as much as possible, and provide an internet based working environment. Therefore, we proposed Astrophysical Integrated Research Environment plan, abbreviated as AIRE.

Generally, AIRE consists of three parts: a Data Archive Center (DAC), a web-based Data Processing Center (DPC), and a Concurrent Research System (CRS). In the following of this paper, we will describe them in detail.

2 Data Archive Center

At present, most astrophysical data are available online. They can be accessed individually via http or ftp. Users can also query these data by integrated software like Browse

or AstroBrowse. However, due to the limited bandwidth of Internet, it is wise to save some frequently used data locally. Thus, the main aim of Data Archive Center is to collect and manage astrophysical data.

The first thing we should do for DAC is to buy computers to accommodate astrophysical data. We have a SGI1450 server in Tsinghua Center for Astrophysics (THCA) now. Because commercial servers are quite expensive, aftertime we prefer to use PC clusters which are much cheaper, easy to upgrade and expand. Next is to install corresponding software. DAC use Linux operating system, MySQL database and Apache web server which are all open-source software, and free too. In order to download and manage data automatically, we will use AstroBrowse which was developed by HEASARC (*http://www.heasarc.nasa.gov/*) and construct local data manage system. Such system is under debugging now. Finally, we need to fill the DAC with astrophyscial data. Till now, we have collected about 600GB data in our server. Because the Tsinghua Center for Astrophysics concentrates on high energy astrophysics, most of data now are in X-ray and γ-ray bands. In future, we want to collect data in other wavebands too. The current status of DAC is listed in Table

Table 1 Current status of Data Archive Center

Hardware	SGI1450 server with 4 PIII700 CPUs, 1TB raid disk, 4GB memory. PC cluster with 8 PCs, 16 PIII 1GHz CPUs, about 4TB raid disk, 4GB memory.(♣)
Software	Linux + MySQL + Apache Web Server Local data manage system. (♣) Astrobrowse (developed by HEASARC) (♣)
Data	ASCA, ROSAT, RXTE, BATSE... ($\sim$ 600GB) Chandra, XMM, DSS ... (♣) data in optical, radio, infrared band (♣)

NOTE: ♣ means that part is under construction.

3 Data Processing Center

Normally, an astrophysicist downloads data and the corresponding software to his personal computer. It always costs him a lot of time and effort to maintain his personal data processing system. If both data and processing software are in a central server, things will be simple. AIRE will construct such kind of Data Processing Center (DPC). In the central DPC server, we will install routine astrophysical software like HEASOFT, CIAO, AIPS, DIFMAP, IRAF, IDL, MATLAB, OCTAVE etc. What an astrophysicist need is just a web browser and access to the internet. He can use the above software, and process the data in central server interactively in IE or Netscape.

Web based DPC has several advantages. Firstly, users do not need to transfer huge raw data and install processing software. What they need to transfer is the intermediate and final results with amount of data much less than original data. Secondly, users from different places around the world can access all the resoures in DPC. If necessary they can pause their work in DPC, and then continue it later, even at other places. DPC will keep the previous working environment. Thirdly, international collaboraton is practical in DPC. For example, if three astrophysicists, whose research fields are in radio, optical and x-ray band respectively, want to study the QPO properties of AGNs, then radio astrophysicist can process the radio data in DPC, and optical and x-ray astrophysicists process the optical and x-ray data too. Finally, they can compare and disscuss their results, and write their scientific papers in DPC.

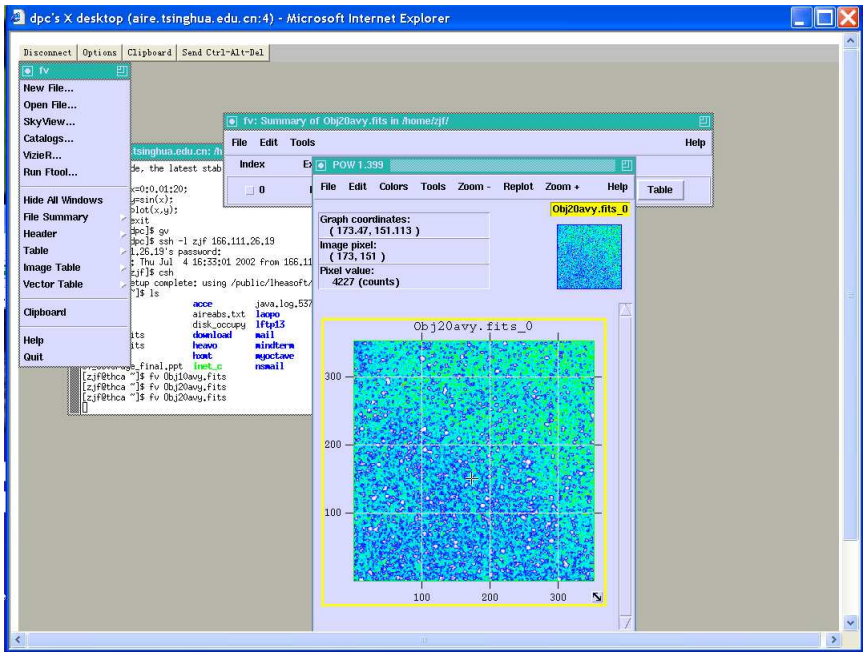

Fig. 1 An example of web-based Data Processing Center

Currently, we have finished a tentative model of DPC (see Fig.1). It will be improved for real use soon later.

4 Concurrent Research System

Astrophysical research associates tightly with computers. Computers were used to control telescopes, and collect the data of objects. Special computer software is also needed to process the observed data, and analyse the results. Even for theoretical astrophysicists, computers are essential. They need software like *Mathematica* to do some complex mathematical dedution.

Even with the help of computer, the power of one astrophysicist is very limited. He/she may only master the background knowledge, observation and data reduction techniques in one or two wave bands. It is difficult for astrophysicists to master all the background knowledge and techniques in many wave bands. Yet our cosmos display its characters in panchromatic bands. Therefore, in order to understand our cosmos further, astrophysicists in different fields should have more effective collaboration.

Owing to the fast progress of internet, now international collaborative research is more convenient than before. For example, CVS (concurrent version system) has been widely and successfully applied in Linux development. Under the control of CVS, computer scientists from different places can explore the same software, although they are geometrically apart. CVS acts like an excellent manager in collaboration.

In astrophysical research, now we also need CVS-like system. So we proposed Concurrent Research System (CRS). CRS will take advantage of CVS technique. It can concert the research work of astrophysicists from different fields as well as from different places. Out of

question, CRS will play an important role in the new discoveries of astrophysics.

5 Discussion

AIRE is in its early development stage. We intend to finish a tentative version in summer 2003. Due to the security reasons, now our servers are only accessible inside Tsinghua university for test. After the tentative version is finished, they will be open internationally.

This work is supported by the Special Funds for Major State Basic Research Projects of China.

References

Boroson T. et al., in ASP Conf. Ser., Virtual Observatories of the Future, Brunner R.J., Djorgovski S.G., Szalay A.S., eds., 2001, 225, 353

Boroson T. et al., in ASP Conf. Ser., Virtual Observatories of the Future, Brunner R.J., Djorgovski S.G., Szalay A.S., eds., 2001, 225, 34