

Hera: Building a Web-Based Analysis Environment at the HEASARC

Christina W. Heikkila¹, William D. Pence, Thomas A. McGlynn², Peter Wilson³, James Peachey¹, Ning Gan¹, Edward Sabol¹, Keith Arnaud⁴

NASA/Goddard Space Flight Center, Greenbelt, MD 20771

Abstract. We are developing protocols which will allow full scientific data analysis to be performed within suitably configured Web browsers. Currently, although the Web is critical to both the data discovery and publishing aspects of the scientific enterprise, it is used far more sparingly in the gritty details of data analysis. The data analysis environment does not share the advantages of immediate access to new resources, quick and convenient downloads, and the support from commercial vendors that data distribution currently enjoys. We believe we can extend existing protocols to provide a complete analysis environment that users will access through their Web browsers. We will implement these protocols within the HEASARC's FTOOLS and XANADU analysis environments to provide a fully functional system to users. We anticipate these protocols would be extensible to other software environments.

1. Introduction

Hera is designed to bring data analysis to the web. It breaks down artificial barriers between data analysis tools and the data archive, between user data and archive data, between resources at the data center and resources on the user's machine. Currently, users of the HEASARC FTOOLS and data archive must install a copy of the FTOOLS at their site, download any data needed from the HEASARC archive, then begin analysis on their local machine. Using Hera, the user could actually use FTOOLS on archive data all on the HEASARC servers, without installing the entire suite of FTOOLS, and without downloading the data files from the archive.

The HEASARC's Hera program has just begun in the last six months, and has been awarded AISR funding for the next three years. We have been exploring the enabling technology and have developed an initial prototype.

¹Emergent IT

²Universities Space Research Association

³Apple Computer, Inc.

⁴University of Maryland

2. Goals

Hera should allow seamless use of software and data resources at the user's local site and the archive. Ideally, the user should not care whose machine is actually running a task, and the user's space on the HEASARC servers should seem like just another local disk.

We plan to offer a scripting capability, where users can put together sequences of analysis tools to be repeated. These scripts will allow a mix of local and server tools. Using this scripting capability, Hera will provide simple "cookbook" tools to allow astronomers (or even expert high-school students) to perform routine data analysis quickly.

We will use Hera as a real service at the HEASARC. Thus we have made implementation choices (e.g., Tcl for the user interface) that reflect current HEASARC architecture. However the protocols and concepts used within Hera should be portable to other systems and languages.

3. Examples of Potential Hera Use

The Integral mission anticipates that response matrices required to do spectral analysis of observations from some instruments may be several gigabytes in size. Either users must clog the web downloading these enormous files, or build a copy of the extensive Integral packages needed to generate these files—with no guarantee that they will run on the user's machine. Using Hera, users will be able to analyze archival or even their own proprietary data without placing such a burden on their local machines.

Hera makes it possible to do real data mining in the HEASARC archives. The user has access to both the full archive and the full software system. Tasks can be run over large numbers of retrieved datasets. For example, a user looking at faint variable X-ray sources might first use the Web to query the HEASARC catalogs for long observations by ROSAT or ASCA. He or she could then use Hera to run a script on each of the resulting observations and download the results to a local machine.

An optical astronomer unfamiliar with high-energy analysis can use Hera scripts to guide her analysis. Using Hera, existing HTML cookbooks can not just tell her what tasks to run, but start them as based on simple web form selections.

With a little care, the previous example can be extended to build educational pages which allow college or even high-school students to interact dynamically with real data in the HEASARC archive.

4. Current Status

Using Hera, users can initiate analysis tools on the HEASARC servers or on their local machine (currently limited). The FTOOLS are presented in a set of hierarchical menus. Choosing one brings up a window where the user can set parameters, including the input and output files. A new window is created showing any text output of the FTOOL, and a graphical window shows any images or plots generated.

Every user is given their own personal space on the HEASARC servers, where FTOOLS results files, status files, and preference files are stored. Symbolic links are used as pointers to files in the HEASARC archive which the user has marked for later analysis. Users can also transfer data files to and from their personal space at HEASARC. Hera provides users with a full set of file management tools to manage their user space at HEASARC.

5. Future

Despite the progress we have made so far, there is still much work to be done. In particular, we do not yet have scripting capabilities available—these will allow a user to string together a series of tools to be run on the same data or on multiple data files. We plan to create cookbook recipes as Hera scripts so users can easily analyze large sets of observations. These scripts would also be valuable as teaching aids—they could be integrated in a college or high school exercise involving analyzing data.

We need to create a system of metadata, so that Hera will know which FTOOLS can accept as input which types of FITS files. The FTOOLS use a parameter file which gives some limited metadata for each tool, but describing HEASARC software and data sets is a substantial element of the work needed for Hera. XML looks like a natural candidate for handling this metadata.

The FTOOLS should be able to recognize that the versions of tools installed on the user's machine are out of date, and offer to download newer versions. Some users may choose not to install a local copy of the FTOOLS at all, and simply run all analysis on the HEASARC servers.

We plan to make the first public release of the Hera system by the end of 2001. We welcome suggestions and comments—please contact:

- Bill Pence : pence@milkyway.gsfc.nasa.gov, or
- Tom McGlynn : tam@lheapop.gsfc.nasa.gov.