

A Data Inventory Service for the Virtual Observatory

Thomas McGlynn, Jeongin Lee¹

NASA Goddard Space Flight Center, Greenbelt, MD 20771

Robert Hanisch, Gretchen Greene

Space Telescope Science Institute, Baltimore, MD 21218

William O'Mullane

Johns Hopkins University, Baltimore, MD 21218

Abstract. We describe a Data Inventory Service (DIS) that uses Virtual Observatory protocols to discover, organize and present astronomers with a rapid precis of what is known about a given region of the sky. The DIS queries a dynamic registry of VO resources to discover potential sources of information, queries them on behalf of the user and then saves and organizes the results for display.

In addition to providing a quick mechanism for getting information about a region, the DIS also provides a mechanism through which these data can be sent directly into analysis tools or downloaded to the users machine.

When a data provider registers a new data resource in a data registry, the DIS will find it automatically for any subsequent requests. Since the DIS caches results from services, users can quickly browse catalog and image results. Cached results are maintained for several days, so when a region of the sky is of particular interest, e.g., immediately after a gamma-ray burst, the cache results may be available immediately to all, not just the first user to query for them.

The DIS is built on top of early protocols developed as part of the US National Virtual Observatory effort, the Cone search and Simple Image Access protocols, and uses the VOTable format extensively. As additional and more sophisticated protocols become available, the DIS will become a more powerful interface to distributed archives and catalogs.

1. Introduction to the Data Inventory Service

The new Virtual Observatory Data Inventory Service (DIS) allows users to quickly find out what is known about a given patch of the sky. While Virtual Observatory Protocols enable scientists to access information from many

¹Universities Space Research Association, Seabrook, MD 20771

diverse sources, the data are useless unless scientists can discover and understand the resources available to them. The DIS provides scientists with a quick and convenient summary of many of the data sources available. The DIS currently shows the users images, observations and cataloged objects in the requested region. Since the DIS caches the results from the various services, users can conveniently browse the images and catalog entries. They can choose to download data of interest, or to initiate analysis tasks in VO enabled environments. Currently these include Aladin and OASIS. The DIS builds upon earlier efforts to provide broad access to astronomy resources, but the combination of catalog and archive access, multi-institutional support, coherent display of results, flexibility in access services, and caching of results give it unprecedented power and convenience.

Data providers only need to register new services in a VO Registry service to make them immediately visible through the DIS (see Greene et al., 2004). Data providers do not need to even know about the existence of the DIS.

Internally the DIS leverages many of the emerging VO standards. Systems are queried using standard Cone Search and SIA protocols, data are returned using VOTables, and these same tables (along with FITS files) are sent to initiate analysis tasks.

The VO DIS is currently available at <http://heasarc.gsfc.nasa.gov/vo>.

2. Implementation

The DIS is a relatively simple tool which fleshes out a framework of emerging VO standards. These include

VO Resource Data Model: A common description for how remote resources are described in the VO.

VOTable: A standard format for tabular (or more complex) data that is easily generated, transported and parsed.

SIAP: The Simple Image Access Protocol provides a standard interface to image data and archives of images.

Cone Search Protocol: A early simple protocol for positional queries of catalogs.

FITS: The VO uses the pre-existing FITS standards for transporting astronomical images and other data. The figure shows the typical flow of data and control in the DIS and the standards involved at each step.

3. Future Plans

As the sophistication of VO protocols grows and as they are implemented by more and more data providers, we anticipate that the DIS will become an increasingly powerful tool for browsing and extracting data from multiple sites. During the next year, we hope to add services that support the new Simple Spectral Image Access Protocol. Standardized data set identifications are being

looked at by several major observatories and these can be used by the DIS to link observation catalogs to observational data. The current links from the DIS to image archives typically point to only a single representative image, but will be enhanced to provide access to all relevant data.

Many more services will be available by the end of the year, and the DIS will need controls that enable users to focus on the data of interest. The DIS will always be usable given just a position, but users may wish to concentrate on data in a given wavelength regime, or on catalogs of a particular class of source. As the number of sources available becomes very large, the DIS' caching strategy will be revised to ensure that it only queries services for data that is likely to be used rather than querying multitudes of sources that are not of interest. This can confuse the user and waste computer cycles and bandwidth.

Currently the DIS can only send data to the OASIS and Aladin services. As the standards for VO services evolve and more of them come on-line, a large number of services will be available. The service selection elements of the DIS will be organized to give the user an overview of the available analysis capabilities. The DIS will serve as a simple tool to enable users to select the data they wish from the multitude of data resources, and pipe it into one or more of the data analysis services that are coming on-line in the VO.

Acknowledgments. This effort was supported in part by NSF Cooperative Agreement AST 0122449.

References

Greene, G., et al. 2004, this volume, 285

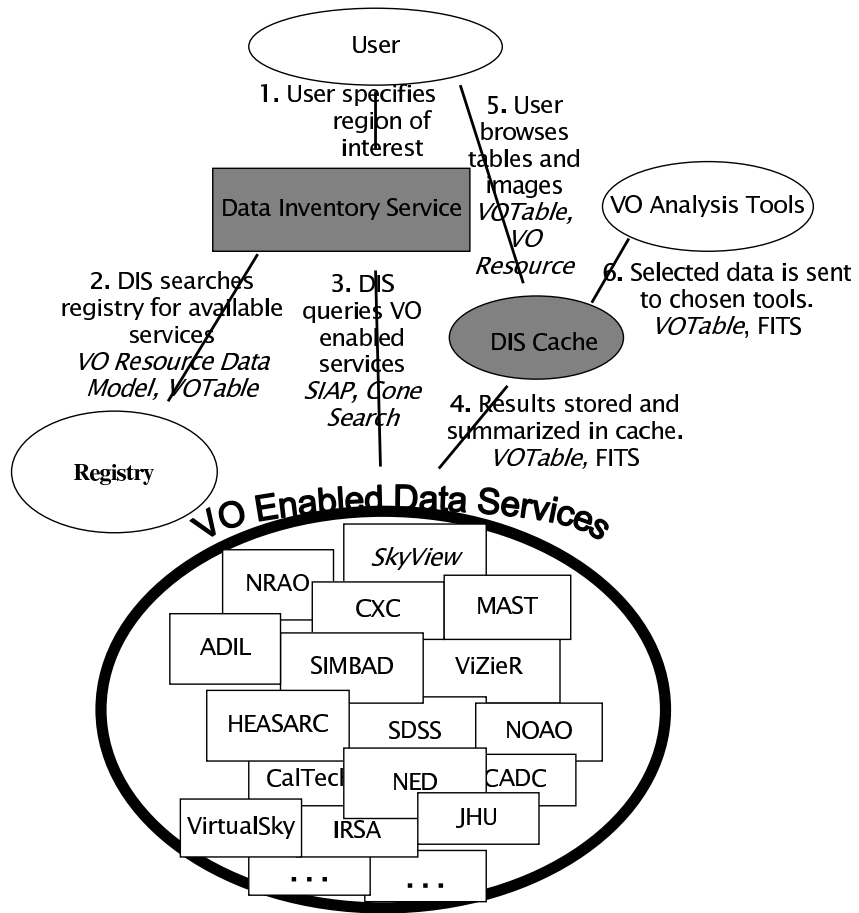


Figure 1. Data Flows in the Data Inventory Service