Virtual Observatory Capabilities of NASA Astronomy Archives

Thomas McGlynn, HEASARC; Gretchen Greene, MAST; Vandana Desai, IRSA/IPAC; Joseph Mazzarella, NED

Version 2.0. February 10, 2015

This report describes the Virtual Observatory (VO) capabilities and practice of NASA astronomical archives as of January 1, 2015. It is provided as part of the NASA Astronomical Virtual Observatories effort and is the first deliverable in the design and implementation of a standard model for VO access by the participating NASA archives. This report describes services at the Mikulski Archive at Space Telescope (MAST), the Infrared Science Archive (IRSA) and NASA Extragalactic Database (NED) at the Infrared Processing and Analysis Center (IPAC), and the High Energy Astrophysics Science Archive Research Center (HEASARC) at the Goddard Space Flight Center (GSFC). Where feasible we have also included discussion of resources at the Astrophysics Data System (ADS) and Chandra Science Archive (CXC) for complete coverage of NASA archives, but these centers are not funded elements of the NAVO collaboration.

This document strives to be both synthetic – providing an integrated view of access to NASA resources through the VO – and complete – detailing the all capabilities available to the community through each archive's services. The intended readership of this document includes the participating archives and NASA management who will use it to develop plans for and assess progress towards our standard archive model. However since we believe that much of the information given here is also of interest to the astronomical research community, we have developed a public facing web site [xxx] that organizes elements of this document in a fashion more appropriate to a researcher's needs. A copy of this document is available on that web site. This document will occasionally refer to pages on the web site which is intended to provide a complete and up-to-date summary for the research community of how to access NASA data through the virtual observatory.

This document provides a comprehensive discussion of mechanisms that enable the community to retrieve NASA (and some non-NASA) datasets from our archive and the registry and other facilities that enable users to discover datasets and services of interest.

1. Introduction

Figure 1 illustrates different kinds of NASA astronomy archives.

During mission operations, the **mission archive** provides access to that mission's data. As the mission nears completion, responsibility for the data passes to the appropriate **domain archive**. Even during the mission there will be very considerable collaboration between the mission archive and the domain archive and functions may be delegated even during mission operations (e.g., between Spitzer and IRSA, or Fermi and the HEASARC). Currently the CXC is the only mission archive actively providing VO services to the community.

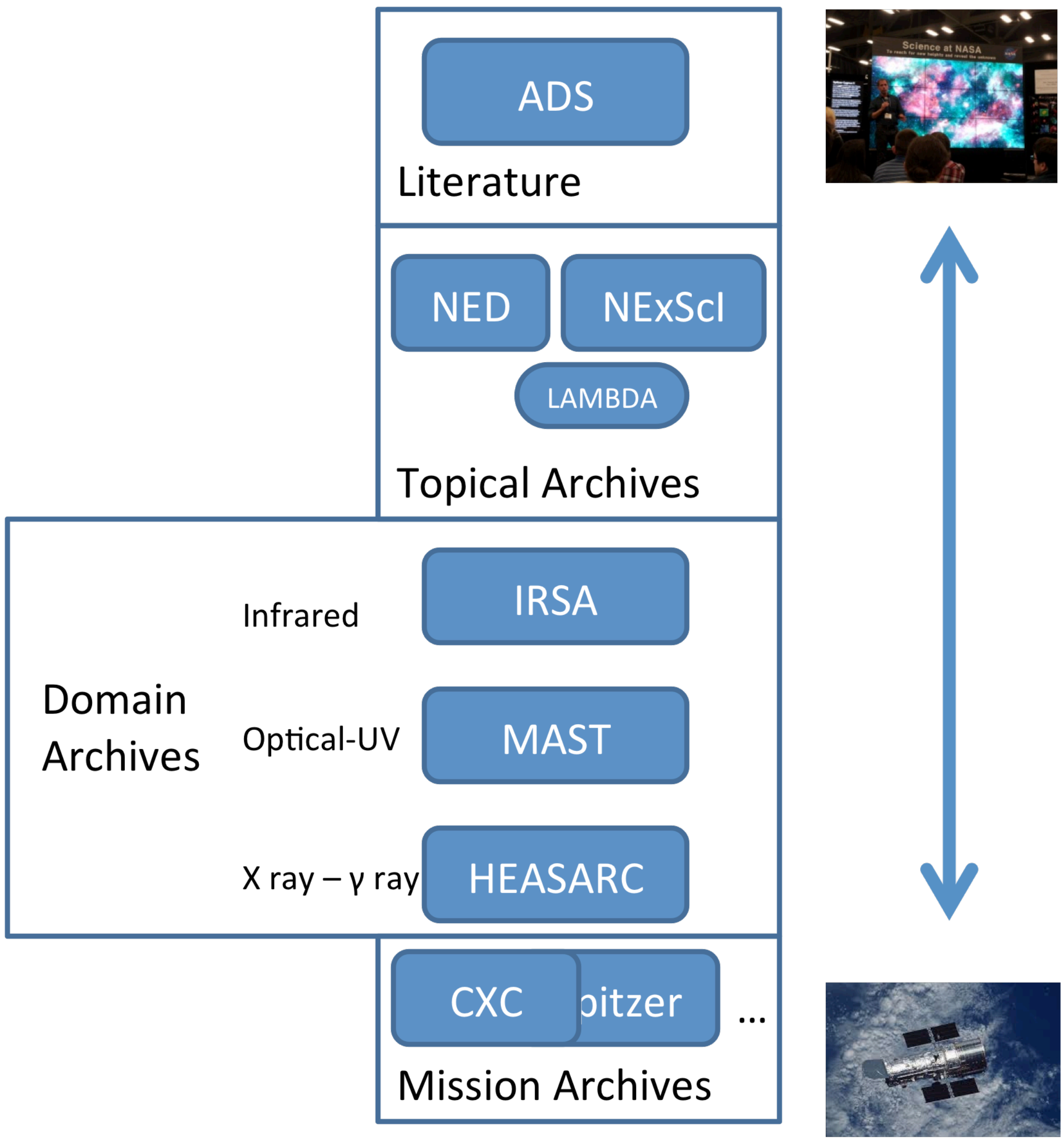

Figure 1. Organization of NASA Astronomy Archives

The domain archives are the core element that provides for the long-term preservation of the mission data and all of the software, documentation and other elements needed to keep the mission data useful. They may supplement NASA datasets with complementary resources from ground observations or non-NASA missions. At the boundaries between the domain archives, there may be shared responsibilities in ensuring that a given mission's data is accessible to multiple groups within the astronomical community.

Topical archives use mission data to provide comprehensive support in specific astronomical areas. NED, NEXScI, and the LAMBDA archive within the HEASARC support extragalactic, extrasolar planetary and CMB astronomy respectively and integrate data from many NASA and non-NASA datasets. While NASA mission data is critical to topical archives, they are organized around the science theme: extragalactic objects, exoplanets, or cosmological parameters so that there may not be sharp boundaries between their NASA and non-NASA holdings.

The ultimate goal measure of success of the mission is the publication of scientific results, for which the ADS is now the primary access point for the astronomy community. While the ADS is primarily a portal to the **literature**, its links back to NASA (and non-NASA) data sources used within published articles ensures the ADS is a powerful archive portal as well. The ADS's coupling of science results and data also provides a clear record of how mission data can and has been analyzed and directly supports NASA's mission infrastructure.

NASA archives from each of these types provide roughly 1,000 distinct registered VO capabilities to enable users to access astronomy data. These services implement one or more of the VO standards described in appendix A. Subsequent sections describe the specific VO access capabilities provided by NASA archives, how NASA VO services are published in the VO registry and made available for user discovery, and a quantitative review of the current use of VO protocols.

2. VO Access to NASA archive data.

We break down access using current VO protocols for table access and data access into categories described below.

Table1 summarizes VO access to NASA mission and project data at all of its astronomy archives. The first column gives the name of the mission/project. The second column gives the spectral regime or regimes associated with the mission data. The primary archive indicates the NASA archive primarily responsible for the mission. In a few cases there are multiple primary archives when different archives have primary responsibility for different aspects of the mission.

The NASA role indicates how/whether the mission/project was funded by NASA

P – NASA was one of the primary sponsors of the mission.

J – NASA is a major partner.

A – indicates that the NASA involvement is primarily in the archive/analysis phase and not in the data collection.

Blank – Limited NASA-funded involvement prior to incorporation of the data in one of the archives. In this case the mission name is shown in italics.

The next three columns describe the table data products that are available in three categories: *Obs* indicates an observation table describing observations taken by the instrument or mission, *Obj* indicates that there is a table of the objects seen by the mission. This is restricted to tables that encompass all or a significant fraction of the mission. The *Science* column indicates that tables associated with specific papers or science results for the mission are available. This might include detailed analysis of objects seen in specific observations, particular classes of objects, correlations with other instruments or other high level science tables. For some missions, e.g., ROSAT, Chandra and XMM there are dozens of these tables available.

The letters in the columns indicate the archives at which the data is available

- A – ADS (Cone search)
- C – Chandra Science Center (Cone and TAP protocols)
- H – HEASARC (Cone and TAP protocols)
- I – IRSA (Cone and TAP [after 3/2015])
- M – MAST (Cone)

Sometimes a table is available from multiple archives

The next three columns indicate the existence of image services (using the SIA protocol) at the institutions. Three kinds of image services are distinguished. *Obs* services return images corresponding to complete observations. *Obj* services return images associated with specific objects or science objectives, e.g., the SIA service that supports the HST M31 mosaic. *Cutout* services provide more or less arbitrary cutouts of the observational data at a user specified location and size. In many cases these return pixel subsets of what would be seen in a full observation cutout, but in others the service may automatically mosaic and resample the base data.

The final column indicates if there is a SSA services associated with the project/mission/instrument. Generally these return the spectrum associated with a single object. NED uses the SSA protocol to return a spectral energy distribution for the object that will merge observations from a myriad sources.

Table 1. VO Access to NASA and Supplementary Data

Mission	**Regime**	**Primary Archive**	**NASA role**	**Tables**			**Images**			**Spectra**
				Obs	**Obj**	**Science**	**Obs**	**Obj**	**Cutout**	**Obj**
2MASS	IR	I	P	I	I		I		I,H	
ACT	Microwave	H								
AGILE	X ray	H			H					
AKARI	IR	I			I,H					
Ariel V	X ray	H		H	H					
ASCA	X ray	H	J	H	H	H				
Astronomy Literature	All	A	A			A				
BBXRT	X ray	H	P	H						

BEFS	UV – EUV	M								M
BeppoSAX	X ray	H		H	H					
BLAST	Sub-MM	I	J	I			I			
BOLOCAM	mm	I					I			
Chandra	X ray	C	P	C,H	C	H	C	C		C
COBE	CMB/IR	H/I	P		H				H	
Compton	ϒ ray	H	P	H	H	H				
Copernicus	UV, X ray	M	P	H						
COS-B	ϒ ray	H		H						
COSMOS	All	I	J		I					
DENIS	IR/Optical	I		I						
DSS	Optical/IR	M	A	M	M				M,H	
Einstein	X ray	H	P	H	H	H				
EPOCh	Optical/IR	M	P							
EUVE	EUV	M/H	P	H	H				H	M
EXOSAT	X ray	H		H	H					
Extragalactic Sources	All	N	A	N	N	N	N		N	N
Fermi	ϒ ray	H	P	H	H				H	
FUSE	UV	M	P	M,H						M
GALEX	UV	M	P	H	M	H			M,H	M
GINGA	X ray	H		H	H					
HEAO-1	X ray	H	P	H	H				H	
Herschel	IR	I	J	H		I		I	I,H	
HETE-2	X,ϒ ray	H	P	H	H					
HLA	IR, Optical,UV	M	P						M	
HPOL	Optical/IR	M								M
HST	IR/Optical/UV	M	P	M,H	M	M,H	M	M	M(many)	M
HST:ACS		M								
HST:COS		M								
HST:FGS		M								
HST:FOC		M								
HST:FOS		M								M
HST:GHRS		M								M
HST:HSP		M								
HST:NICMOS		M								
HST:STIS		M								M
HST:WF/PC1		M								
HST:WFC3		M								
HST:WFPC2		M								
HSTONLINE		M								
HST (COSMOS)									M	
HST (GOODS)		M					M		M,H	
HUT	UV	M	P	M						M
IMAPS	UV	M	P							
INTEGRAL	X,ϒ ray	H	J*	H	H					
IRAS	IR	I/H	P	I					I,H	
IRTS	IR	I					I			
ISO	IR	I		H	H					

IUE	UV	M	P	H						M
K2	Optical	M	P							
Kepler	Optical	M	P	M	M	M				
MAXI	X ray	H			H					
MSX	IR	I			I,H		I			
NuSTAR	X ray	H	P	H						
Planck	mm/IR	I	J		I,H				H	
PPMXL	IR, Optical	I			I					
PTF	Optical	I								
ROSAT	X ray, EUV	H	J	H	H	H			H	
RXTE	X ray	H	P	H	H				H	
SAS-2	ϒ ray	H	P	H						
Spitzer	IR	I	P	I,H	I	I,H	I	I(>10)		
SPT	Sub-MM	H								
Suzaku	X ray	H	J	H						
SWAS	IR	I/H	P							
Swift	X,ϒ ray, UV/Opt	H	P	H		H			H, M	
TUES	UV	M	P							M
UIT	UV	M	P				M			
USNO	Optical	I			I					
Vela 5B	ϒ ray	H								
VLA-FIRST	Radio	M					M		M,H	
WISE	IR	I	P		I		I		I,H	
WMAP	CMB/IR	H/I	P		H					
WUPPE	UV	M	P							M
XMM-Newton	X-ray, UV/Opt	H	J	H	H	H			M(OM)	

2.1 Tables

Observation Tables

For most missions, the mission data is divided in temporally bounded observations. While the details vary from mission to mission there are typically a time range, exposure, position, target, investigator, and instrument configuration associated with the observation. Observation tables provide a summary of the mission. Users interested in data from a given instrument may query these observation tables to determine data of interest. For some missions, e.g., IRAS there may not be an obvious breakdown into discrete observations, but there may still be standard set of mission products – e.g., the IRAS all sky survey tiles.

All four of the VO data access protocols may be used to query observations: The simple cone search protocol can be used for any observations where the observations have defined positions. The SIA and SSA protocols can be used for imaging and spectroscopic instruments respectively. TAP interfaces can be used for all types of data. The SIA and SSA access will be described below in the data access section.

For most mission datasets with positional information, cone search capabilities are available. TAP services are available for missions hosted at the HEASARC and CXC archives and for a few external missions where an observation table has been mirrored at the HEASARC. Image observations are available for a number of IRSA and MAST hosted missions, while spectra are available for many spectral instruments hosted at MAST and for Chandra data through the CXC. The detailed standardized information provided by the initial response in SIA and SSA requests can be useful to astronomers even if the user does not download the files pointed to be the returned table. I.e., since more metadata is defined in the SIA/SSA responses than in the Cone Search response, tools may be able to burrow deeper without require as much manual guidance.

Object tables

For many missions a major data product is an object catalog (or catalogs) derived from systematic analysis of the observations. For many science analyses, the astronomer can use the information in these catalogs without needing to reference the observations directly. Such catalogs are typical in imaging missions with reasonably high resolution. For something like a narrow-field spectrometer, or a simple collimated X-ray detector, the observation catalog may serve dual duty and include both observational constraints and measured values. Some missions may have multiple object catalogs reflecting different aspects of the mission. For example, ROSAT has distinct object catalogs from its all-sky scanning phase and the pointed observations. Object catalogs for a full mission may be difficult to construct especially for mission like HST where the data is extremely heterogeneous with many different instruments, filters and modes. Given their usefulness, an HST object catalog is being created despite these difficulties. The relationship of the data and object catalogs can sometimes be obscured by terminology. E.g., one can consider the Guide Star Catalog 2 to be the object catalog associated with the Digital Sky Survey datasets. Generally where mission object catalogs have been produced, the NASA archives provide cone search interfaces to them. Object catalogs at the HEASARC and CXC are also accessible through TAP.

Analysis tables

A third category of tables includes those that reflect analysis of various samples. These tables often include the correlation of data between different missions, or between mission and ground datasets or they may define some specific class of objects. The HEASARC and IRSA provide cone search access to over 1000 such such catalogs. Those available at the HEASARC can also be queried using TAP.

NED's cone search interface provides a unique capability to look for extragalactic data in a given region of the sky globally across all types of astronomical observations. The ADS cone search returns papers which discuss objects in a given region. While this potentially points to a very refined scientific results, in practice NED's association of data into objects is more immediately useful.

The VO protocols define the structure of URLs for querying the archive relative to some base URL. Although there are roughly a thousand table querying services available through the NASA archives most of these use base URLs which a standardized variations on roughly a dozen basic structures described in Table 2. This table also notes the limits associated with various services. [Needs to be finished.]

Table 2. Table Access VO Services at NAVO archives.

Service	P	N	Limits	Base URL
ADS	Cone	1	7,8	adsabs.harvard.edu/cgi-bin/abs_connect?data_type=VOTABLE&
MAST/STScI	Cone	12		archive.stsci.edu/XXX/search.php?
MAST/Kepler	Cone	3		archive.stsci.edu/kepler/XXX/search.php?
MAST HST instruments	Cone	9		archive.stsci.edu/hst.php?sci_data_set_name=X*&
??	Cone	2		archive.stsci.edu/spec_class/search[2].php?
MAST GALEX	Cone	1		galex.stsci.edu/gxWS/ConeSearch/gxConeSearch.aspx?
CXC Chandra	Cone	2		cda.harvard.edu/XXX/coneSearch?
CXC Chandra	TAP	1		cda.harvard.edu/cxctap
HEASARC Tables	Cone	733	1,2	heasarc.gsfc.nasa.gov/cgi-bin/vo/cone/coneGet.pl?table=XXXX&
HEASARC Tables	TAP	1	1,2,3	heasarc.gsfc.nasa.gov/xamin/vo/tap
IRSA 2MASS	Cone	2		irsa.ipac.caltech.edu/cgi-bin/Oasis/CatSearch/npa-catsearch?cat=fp_XXX&
IRSA Tables	Cone	385	5,6,9	irsa.ipac.caltech.edu/SCS?table=XXXX&
NED	Cone	1		ned.ipac.caltech.edu/cgi-bin/NEDobjsearch?search_type=Near+Position+Search&of=xml_main&
Total		**1153**		

Limits:

1. No limits on size of query
2. 100,000 row limit on results
3. Result limit can be overridden
4. Services not indexed in registry
5. 0.75 degree radius limit on query
6. 50,000 row limit on results
7. Echoes input position in results
8. 200 row limit on results
9. 25 currently in registry

Most table services have at least a default limit on the number of rows that will be returned in a query. In some cases this can be overridden by the user (there is a standard way to do this in TAP). Other services additional constraint the radius of cone search requests.

2.2 Data Access

Observation Data

For most of their imaging missions MAST and IRSA provide SIA services that enable users to retrieve observation images. These are the actual images generated by the observations at various levels of processing. Associated calibration images are sometimes included. For some missions, e.g., Planck and IRAS, the basic images are not taken as discrete observations but are generated in post processing. However there is still a standard set of mission images to which SIA services provide access.

MAST provides SSA access to spectral data for spectrometer based missions and the for the HST spectrometer instruments.

Derived Images and Spectra

MAST, IRSA and the HEASARC provide many images services for specific regions of the sky or associated with specific objects. Certain regions of the sky – notably the GOODS and COSMOS regions have many specialized services associated with them. These specialized SIA services are noted in the SIA column. They typically do not provide access to the full mission dataset but are often more highly processed (e.g., multiple observations may be mosaicked together).

Other SIA service are direct cut-out capabilities where the user can specify a particular region of interest and will get results if that area is in the mission coverage region. IRSA provides a standard cutout service for its imaging missions. MAST provides cutouts for several key datasets, notably GALEX, the Digitized Sky Surveys and the FIRST data. At the HEASARC, the *SkyView* service provides cutout capabilities for over 100 sky datasets. This includes data from 17 NASA missions/datasets as well as many ground and non-NASA datasets. *SkyView* can be invoked as a generic SIA service that returns all surveys it knows of or through specialized services that query specific sets of surveys (e.g., all of the Fermi or SDSS wavebands).

NED's SIA service provides a combination of cutouts and observations associated with specific objects. This can include highly processed images as well as lower level data.

Table 3 describes the basic patterns used in data access VO protocol implementations at NASA archives and notes the limits of these services.

Table 3. Data Access VO Services at NAVO archives

Services	P*	N	Limits	Base URL
MAST GOODS	i	1	3,4	http://archive.stsci.edu/eidol.php?
MAST Image Services	i	36	3,4	http://archive.stsci.edu/siap/search.php?id=*XXX*&
MAST Spectra Services	s	11	3,4	http://archive.stsci.edu/ssap/search2.php?id=*XXX*&
MAST GALEX Spectra	s	1	3,4	http://galex.stsci.edu/gxWS/SSAP/gxSSAP.aspx?
MAST Scrapbook	i	1	3,4	http://archive.stsci.edu/siap/search.php?representative=y&
MAST EPO	i	2	3,4	http://archive.stsci.edu/stpr/vo_search.php?
MAST DSS	i	4		http://chart.stsci.edu/gscvo/DSS*X*.jsp?
MAST GALEX	i	1		http://galex.stsci.edu/gxWS/SIAP/gxSIAP.aspx?
MAST PR	i	1		http://hubblesite.org/cgi-bin/sia/hst_pr_sia.pl?
MAST Hubble Legacy Archive	i	1		http://hla.stsci.edu/cgi-bin/hlaSIAP.cgi?imagetype=best& inst=ACS,ACSGrism,WFC3,WFPC2,NICMOS,NICGRISM,COS,STIS,FOS,GHRS& proprietary=false&
IRSA Atlas	i	59	1	http://irsa.ipac.caltech.edu/cgi-bin/Atlas/nph-atlas?mission=*XXX*& hdr_location=*YYY*&SIAP_ACTIVE=1&collection_desc=*ZZZ*&

IRSA AllWISE	i	1		http://irsa.ipac.caltech.edu/ibe/sia/wise/allwise/p3am_cdd
IRSA WISE AllSky	i	5		http://irsa.ipac.caltech.edu/ibe/sia/wise/allsky/*XXXX*? http://irsa.ipac.caltech.edu/ibe/sia/wise/cryo_3band/*XXXX*? http://irsa.ipac.caltech.edu/ibe/sia/wise/postcryo/2band_p1bm_frm?
IRSA WISE Prelim	I	3		http://irsa.ipac.caltech.edu/ibe/sia/wise/prelim/XXX http://irsa.ipac.caltech.edu/ibe/sia/wise/prelim_postcryo/2band_p1bm_frm/xxxx
IRSA Merged WISE	I	2		http://irsa.ipac.caltech.edu/ibe/sia/wise/merge/XXXX
IRSA 2MASS	I	5		http//sha.ipac.caltech.edu/ibe/sia/twomass/XXX/XXXX
IRSA Spitzer	i	2		http://sha.ipac.caltech.edu/applications/Spitzer/VO/VOServices? SERVICE=SIAP&DATASET=ivo://irsa.ipac/spitzer.level*X*&;
HEASARC *SkyView* General	i	1	2	http://skyview.gsfc.nasa.gov/cgi-bin/vo/sia.pl?
HEASARC *SkyView* Survey	i	48	2	http://skyview.gsfc.nasa.gov/cgi-bin/vo/sia.pl?survey=*XXX*&
NED SIA	i	1		http://irsa.ipac.caltech.edu/cgi-bin/Atlas/nph-nedsiap? mission=NED&hdr_location=\NEDDataPath\&SIAP_ACTIVE=1& collection_desc=NASA/IPAC+Extragalactic+Database+Image+Data+Atlas+(NED)&&
NED SED Service	s	1		http://vo.ned.ipac.caltech.edu/services/querySED?REQUEST=queryData&
CXC Chandra	i	1	/cscsiap/?	http://cda.harvard.edu/cxcsiap/queryImages?
CXC Chandra Grating images	i	1		http://tgcat.mit.edu/tgSia.php?
Total Count		**162**		

*i=SIA, s=SSA

Limits:

1. Variable limit on query size but typically of order 10 degrees.
2. By default results returned depend on ratio of query size and intrinsic pixel resolution of survey datasets.
3. Default limit of 2,000 rows returned
4. Default can be overridden.

2.3 Other VO Access and Integration

In addition to access direct implementation of VO table and data access protocols, NASA archives support a number of other capabilities that enable easy integration of NASA data in the virtual observatory and efficient multimission access to NASA archive data.

MAST's portal services and the HEASARC's standard Xamin query tools support the VO Simple Application Messaging Protocol (SAMP). This means that when users run web queries with these tools the results can be sent directly to other VO-enabled tools.

The standard VOTable format is universally supported as an output format for table queries even when these queries are made through non-VO protocols.

Many non-VO interfaces to NASA archive holdings are discoverable through the VO registry. In particular NED publishes the interface to 17 specialized non-VO capabilities. Some of these, notably the NED name-resolver are fundamental tools used extensively through the community. In this case the very ubiquity of the existing implementation may have obviated the need for a specifically VO protocol.

The VO registry standards allow a given resource to describe multiple interfaces. The HEASARC (and others?) use the registry to describe both VO and custom interfaces for all tables.

MAST uses VO interfaces extensively in the implementation of its portal interfaces. These internal requests are not included in the statistics given below. The HEASARC uses the SIA interfaces to 2MASS, WISE, the SDSS, and the UKIDSS surveys to access data for *SkyView*.

2.4 Legacy VO Science Services

Although not directly supported by the NAVO project, there are significant VO capabilities hosted at NAVO archives that are legacies of earlier VO efforts. The most important of these are the cross-correlation services hosted by IRSA and the VO Portal hosted at MAST (and using HEASARC web-services). No further development of these as particularly VO capabilities is envisaged, but the VO portal capabilities are being subsumed within the ongoing MAST portal efforts. These projects had always had substantial sharing of code and manpower.

3. Finding NASA VO Data Resources

The primary VO mechanisms for publishing and discovering information are the VO registry protocols which are supported in a variety of ways at the different centers.

From the user perspective, the primary NASA VO registry is hosted at MAST which supports a full publishing and queryable registry. This is the only fully searchable VO registry supported in the US and is a critical element of the global VO infrastructure. This system currently supports all the standard features of an IVOA registry including standard interfaces for searching registry contents, a standard relational registry data model, and a publishing service for distributing the registry records. In addition, the MAST VO registry provides a community publishing system. This enables external users to enter new records or edit existing ones through a web interface. The MAST VO registry provides a web portal which is referred to as the Directory. This allows users to search users to search the directory using either structured of keyword queries (see http://vao.stsci.edu/directory/).

The HEASARC and Chandra Science Center support publishing registries. These enable the centers to publish and update records describing the VO capabilities they host. Publishing registries are not generally expected to be queried by users. The query interfaces they support facilitate discovery of new and updated resources, but provide little support for the kinds of queries that users make that involve astronomical criteria. They are periodically polled (typically daily) by the VO queryable registries – notably the MAST registry described above – and any updates are pulled into the queryable registries.

Users may then use the Directory service hosted at MAST or other VO registry query capabilities to discover the resources at the HEASARC and CXC.

Publishing registries are themselves registered in the Registry of Registries (RofR) currently hosted at NCSA. This is nominally where they are discovered by other registries to enable their being polled. In practice new registries appear rarely so that the MAST registry can use a static list of registries to poll which can be manually updated even if the registry of registries is unavailable.

The other NASA centers have heretofore used the publishing capabilities of the MAST registry directly to publish resource descriptions. This works well when only a few services are supported by the center but can be cumbersome when there are many published services. IRSA is currently looking at building their own publishing registry since they are in the process of publishing several hundred object tables. [Is this true?]

3.1 Registry Content

The structure of the registry entries ensures a basic common structure for all described services. For services that implement specific VO protocols this ensures that registry clients can discover the exact URLs needed to make queries. Several critical elements of the registry are free-form text notably the titles and descriptions of the services. Currently these are no IVOA or NASA standards for these fields. Since these descriptions play a major role in how and whether astronomers can discover useful resources a more systematic approach will be part of our common archive interface.

Currently all NED services share a common description. Specialized descriptions are … [Joe?]

While cone search services for IRSA object catalogs are in place for several hundred services, currently only a few of these are registered. Registry entries for all tables will be made available in the next few months [Steve?]

For cone search services the registry allows us to publish a separate description for each searchable table. For TAP services, where hundreds of tables may be queried through a single VO resource, it is less clear how to publish appropriate descriptions of the tables. One of the main uses of the registry by scientists is to do keyword matches to find data of interest. It is currently unclear how best to preserve this capability for aggregated TAP services. It may be that users will find tables using the cone search entries and then query them using TAP.

3.2 Publishing Services for the Community

As noted above the MAST VO registry provides registry publishing services for the community. Any astronomy data provider and add records to the registry to describe the resources they are making available to the community. The web interface provides extensive help and guidance for entering data. If multiple resources are to be described users clone and modify a master resource rather than entering common information multiple times. Users can describe implementations of VO interfaces and also non-VO capabilities.

Once the record has been entered the MAST registry provides these records to other VO registries around the world so that the service is published universally.

Before publishing services in a registry, external institutions must first establish one or more authority ids that will be used to identify their records. Currently the MAST registry has been used to publish records for about 30 distinct authority IDs external to MAST. This includes other NASA centers, US facilities like NOAO or the SDSS, and even some foreign sites like the Australian National University or the Spanish Virtual Observatory.

3.3 Institution VO web descriptions

In addition to the VO registry each of the archives provides summary pages of VO capabilities but these are not currently standardized. One of the goals of this project will be to develop and complete an standardized template describing the VO services at each center. Combined information will be available in our public facing web site.

Table 5. Institutional VO Web pages.

Archive	URLs
IRSA	http://irsa.ipac.caltech.edu/voapi.html
NED	http://vo.ned.ipac.caltech.edu/ http://www.aspbooks.org/publications/382/165.pdf
MAST	http://archive.stsci.edu/vo/
HEASARC	http://heasarc.gsfc.nasa.gov/docs/archive/vo/

4. Updating NASA VO Data Resources

Many of the resources provided by NASA archives are dynamic especially those associated with active mission. In no case is in any special VO specific action required for updates to such resources to be made visible through the VO protocols. They see the same updates that the archives proprietary interfaces do.

However some VO-specific actions may be required when new resources are created or (occasionally) deleted or when the metadata describing the resource is updated. At all three of the domain archives, the VO cone search interfaces are provided as one of the standard query capabilities for our standardized underlying web query interfaces. The release of the standard mission interfaces includes the new VO cone search capability for search the mission observation tables. However updates to the VO registry to include the new service are likely to require one or more manual steps at the HEASARC and IRSA [not sure if this is automated at MAST]. No special action is required for new tables to appear in the HEASARC's TAP interface.

Add new datasets to the data access protocols sometimes requires a little custom work. While the Atlas system at IRSA and the SkyView system at the HEASARC provide an SIA interface automatically for each new entry, MAST's SIA and SSA services typically requires some level of manual selection of the appropriate files to be pointed to by the services. New entries in IRSA must be manually added to the registry. For *SkyView* new surveys are automatically included in the generic SkyView SIA service, but a file must be manually updated to publish a new entry in the VO registry.

5. In Development

Both MAST and IRSA are currently engaged in developing TAP interfaces to their holdings. The MAST effort includes a standardized model for data products that is reasonably compatible with the VO Observation Data Model. A major issue with both services is the handling of the geometry functions described in the VO Astronomy Data Query langauge. The TAP specification allows sites to implement this to varying degrees. The HEASARC's TAP existing implementation supports only a few basic geometry constructs. MAST and IRSA are still investigating the level at which they can provide effective support.

IRSA is currently engaged in providing standardized access to many hundreds of tables – mostly what would have been classified as science tables above.

NED and the HEASARC are looking for substantial improvements to their data access capabilities but these efforts are nascent and will be reviewed as part of the NAVO effort. NED has been looking updating its SED capabilities in light of the final published SSA standard and also incorporating support for version 2 of the SIA protocol. The HEASARC is planning to provide direct access to observational data through SIA and SSA in addition to the current SIA interface to survey data through SkyViewl

6. VO Data Operations at NASA Archives

Most NASA VO capabilities are deployed as part of the general archive services. No special operational support is required for the data access services and these services are maintained along with the other archive interfaces. The two major areas where there is dedicated operational support for the VO are the registry operations at MAST and the VO operations monitoring services hosted at the HEASARC. These services are substantially independent of the non-VO activities at the centers

The registry operations are largely automated but manual intervention is required to address errors in the registry especially when users who created the registry entries are either no longer active or cannot recall their authorization information to fix the errors. Several times a year services that no longer seem to be actively supported are deleted from the registry. The distributed nature of the VO registry sometimes makes it difficult to know where to update the problematic entries.

VO operations monitoring is hosted at the HEASARC which monitors VO service hosts on an hourly basis and validates all registered VO data services monthly. These data are collected into a queryable database which can be accessed publicly. Information of the average uptime of hosts or the current validation status of each registered service is generally available. These automated processes are supervised by HEASARC personnel who notify sites when services go down. Periodic summaries of the validation status of services are also generated and distributed to sites to describe why services are failing validation and how services could be changed to meet the protocol specifications. If a service persistently fails validation to the extent that it appears not to be returning any useful information it is put a list of services that may be deprecated at the MAST registry. The site owner is informed of this candidate status and if no response is received the service is deprecated at MAST.

7. Usage of VO Resources

The usage of VO services varies dramatically from service to service. In some cases VO services are a substantial fraction of the queries made against a system. Providing useful statistics is challenging. Detailed weekly and monthly statistics are now being compiled as part of the NAVO project. Generally we have tried to note each invocation of a VO data access protocol. Table 4 shows typical numbers (averages for December 2014) and also tries to give comparable numbers for non-VO access for similar queries. E.g., when we compare the NED VO SED service with non-VO access, we look only at specific NED SED requests, not all all NED requests, since the volume of name-resolution requests – for which these is no VO equivalent—would dwarfs the data requests. These numbers give the average number of requests daily. Some data are collected weekly and some monthly so that the periods are not identical, but within that constraint these are average numbers for December 2014.

Table 4. Typical Daily Usage of VO services: Average rates for December 2014

	HEASARC		**IRSA**		**MAST**		**NED**	
	VO	**Non-VO**	**VO**	**Non-VO**	**VO**	**Non-VO**	**VO**	**Non-VO**
Tables	15,407	4,538	76	21,206	1,352	NA	565	10,067
Images	2,667/105	14,449	3,502	NA	2,027	NA	In IRSA	25,960
Spectra	NA	NA	NA	NA	1,688	NA	508	2,408
Registry	NA	NA	NA	NA	1,541	NA	NA	NA

For the HEASARC the Table numbers include both the cone search and TAP requests, over 99% of the numbers are from the cone search protocol. For Images where there are two numbers in the VO column are the number of VO SIA requests made, and the number of images generated from such requests. The Non-VO column gives the total number of images generated (including VO requests) from all interfaces. NED SIAP requests are redirected to IRSA.

Appendix A: Virtual Observatory Protocols and Standards Summary

This appendix provides a brief description of the various VO protocols discussed in the document. For further information please consult the IVOA web site: http://ivoa.net.

- *VOTable*: This defines an XML standard for astronomical tables. The structure is similar to that of FITS tables but there are no arbitrary limits on record sizes and headers and a single VOTable file can contain a nested hierarchy of subtables. There are a number of ways of representing the data content but with NASA archives primarily use a simple text based approach. VOTables include standardized metadata for their content include *Uniform Content Descriptors* (UCDs) which describe the semantics of a given column, and *UTypes*, which describe the specific role of a column in some associated data model.
- *Cone Search*: This protocol allows a user to request a VOTable giving the results of a positional search within a specified radius of a single point. The only required output fields are a name and position for each row.
- *Table Access* (TAP): This protocol allows a user to initiate queries against a database using a specialized version of SQL call ADQL (for *Astronomy Data Query Language*) that includes some support for geometry constructs. However implementers are free to choose the level at which geometry constructs are supported. An implementation may support user uploaded tables (which can then be used in a cross-correlation). Data is returned in VOTables.
- *Simple Image Access* (SIA): In this protocol the user searches for images in a box around a given point. The initial response is a VOTable which provides links to the actual images as URLs. There are a number of standard columns which give WCS information on the returned images and any other metadata the implementor chooses may be returned. The implementation can provide access to raw data to cutouts, or to information that has been dynamically generated. Version 2 of the protocol extends support to data cubes and provides for more standard ways to constrain the requested images (e.g., by spectral regime).
- *Simple Spectral Access* (SSA): This protocol is very similar to the SIA in how it is used, but returns links to spectra rather than images. Given the wide variety of spectral formats used, a wide array of metadata columns are supported though relatively few of them are required.
- *Simple Application Messaging Protocol* (SAMP): This protocol is designed to enable a two or more desktop applications (possibly including applications running in a web browser) to communicate.