PROCEEDINGS OF SPIE

Providing comprehensive and consistent access to astronomical observatory archive data: the NASA archive model

Thomas McGlynn, Giuseppina Fabbiano, Alberto Accomazzi, Alan Smale, Richard L. White, et al.

SPIE.

Providing comprehensive and consistent access to astronomical observatory archive data: the NASA archive model

Thomas McGlynn, NASA Goddard Space Flight Ctr. (USA); Guiseppina Fabbiano, Alberto Accomazzi, Smithsonian Astrophysical Observatory (USA); Alan Smale, NASA Goddard Space Flight Center (USA); Richard L. White, Thomas Donaldson, Alessandra Aloisi, Theresa Dower, Space Telescope Science Institute (USA); Joseph M. Mazzerella, Rick Ebert, Olga Pevunova, David Imel, Graham B. Berriman, Harry I. Teplitz, Steve L. Groom, Vandana R. Desai, Walter Landry, Infrared Processing and Analysis Ctr. (USA) and California Institute of Techology (USA)

ABSTRACT

Since the turn of the millennium a constant concern of astronomical archives have begun providing data to the public through standardized protocols unifying data from disparate physical sources and wavebands across the electromagnetic spectrum into an astronomical virtual observatory (VO). In October 2014, NASA began support for the NASA Astronomical Virtual Observatories (NAVO) program to coordinate the efforts of NASA astronomy archives in providing data to users through implementation of protocols agreed within the International Virtual Observatory Alliance (IVOA).   A major goal of the NAVO collaboration has been to step back from a piecemeal implementation of IVOA standards and define what the appropriate presence for the US and NASA astronomy archives in the VO should be. This includes evaluating what optional capabilities in the standards need to be supported, the specific versions of standards that should be used, and returning feedback to the IVOA, to support modifications as needed.

We discuss a standard archive model developed by the NAVO for data archive presence in the virtual observatory built upon a consistent framework of standards defined by the IVOA.   Our standard model provides for discovery of resources through the VO registries, access to observation and object data, downloads of image and spectral data and general access to archival datasets. It defines specific protocol versions, minimum capabilities, and all dependencies. The model will evolve as the capabilities of the virtual observatory and needs of the community change.

Virtual observatory; data archives; standards; IVOA; NASA

1.   INTRODUCTION

For more than a decade the astronomical virtual observatory effort has been defining protocols that enable users to access data from providers throughout the world in nominally standardized ways. A range of standards exist to enable access to images, spectra, tabular data, resource metadata and more. These standards are now sufficiently mature to support many science use cases, but there has been only limited work done in understanding how these standards can be most effectively combined to provide comprehensive access to observatories' resources.

Standards development in the astronomical virtual observatory effort is coordinated by the International Virtual Observatory Alliance (IVOA) a consortium of 20 national organizations. The IVOA establishes priorities and approves recommendations for data formats, models and protocols. There are now about 40 approved recommendations by the IVOA many with multiple versions.

In October 2014 NASA funded the NASA Astronomical Virtual Observatories (NAVO) program to provide continued support for the US Virtual Observatory infrastructure developed under earlier programs: the National Virtual Observatory (NVO) and Virtual Astronomical Observatory (VAO). The four funded organizations in NAVO: the Mikulski Archive at Space Telescope (MAST), the High Energy Astrophysics Science Archive Research Center (HEASARC), the Infrared Science Archive (IRSA) and the NASA Extragalactic Database (NED), also committed to defining and implementing comprehensive virtual observatory access to NASA astronomy data. While each of the centers has long been an active participant in the virtual observatory, NAVO's inception gave us the opportunity to step

back and assess the publication of our resources in the virtual observatory in a comprehensive ways across the various archives' holdings and looking at the entire suite of available VO protocols.

This paper describes this effort. We discuss our goals, the role of the specific VO standards used, the status and schedule for the implementation of this interface, and our plans to evolve to take advantage of new protocols and address new science requirements. The NAVO archives' perspective is one of integration of VO protocols into existing and mature archive systems. This can be more challenging than developing new archives in the virtual observatory since we have a substantial legacy of resources that predate the standards and models developed within the VO.

In this paper we have attempted to minimize the usage of VO jargon. Some terms do creep into the discussion and are indicated with **bolding** in the initial usage.

1.1 History

The formal development of the virtual observatory protocols in astronomy can be traced to conferences at Caltech [1] in 2000 and ESO/Garching [2] in 2002. These conferences articulated a vision of a research environment in which the nominal host of data was of limited import, where astronomers could easily combine information from multiple sources. The discussions here were of the multi-mission and multi-wavelength science. In 2002 the International Virtual Observatory Alliance [3] was founded to provide an international framework for the development of virtual observatory standards. Since its inception, its focus has been on defining standards. Initial standards for data formats, simple positional queries and image queries have been followed by many more data access, metadata and data modeling standards. . Over the years there has been significant effort to work with scientists and client developers in workshops and summer schools creating a substantial literature for using VO protocols and tools (e.g., see [4]). There has been a deliberate effort to guide the IVOA's efforts towards emerging areas in astronomy: most recently toward support for multidimensional and temporal data.

This focus on the needs of the research community, the clients of VO protocols, was not matched by comparable support for the data providers. While major astronomy archives adopted VO protocols as they were developed, there was limited attention to understanding how a mission *should* publish data in the virtual observatory. For many years while VO protocols supported the publication of some high level products handling the kinds of complex datasets that many telescopes – and certainly most NASA missions – generate, was beyond them. VO protocols can require substantial effort to adapt existing data to them. The fluidity of some VO protocols could also be daunting with significant effort made obsolete when standards changed incompatibly. There was little guidance on how a fabric of VO protocols could be woven to provide reasonably comprehensive mission support. These concerns made it difficult for major institutions to develop a comprehensive strategy for data publication in the virtual observatory.

Recently the IVOA has approved protocols that enable support for the distribution of generalized data sets and has begun to focus more on details of extracting information from archives. It appears that there will be few disruptive changes to existing standards in the near term. It is in this context that the NAVO archives began a review of the implementation of VO interfaces to their data. Our goal was to provide comprehensive and consistent access to NASA's astronomy holdings. This assessment looked at approved and pending IVOA standards, the state of existing implementations at the archives, the costs of implementation of new capabilities and the needs of the astronomical community. This paper discusses the results of this assessment, our belief that it is now feasible and appropriate to provide a comprehensive interface to our data through VO protocols in a consistent way across NASA archives, and our progress in implementing this interface.

1.2 Scope and Goals

NASA is one of the premiere organizations supporting research in the world astronomical community and its archive holdings include data ranging over the entire electromagnetic missions. NASA missions and missions where it has partnered with European and Japanese counterparts are foundational elements of the today's astrophysics research. Figure 1 illustrates the diversity of a few of the major datasets held by NASA. NASA mission data are generally supported by three domain archives: The Infrared Science Archive hosted at the Infrared Processing and Analysis Center (IPAC) is NASA's primary archive in the infrared, the Mikulski Archive at Space Telescope (MAST) is the primary archive in the optical and near UV, and the High Energy Astrophysics Science Archive Research Center

(HEASARC) at NASA/GSFC is the primary archive for X-ray and gamma-ray data. To enable the kinds of multi-mission, multi-wavelength studies that are the basis of much current research these NASA archives also host some key ground datasets, notably the FIRST radio survey, and soon the Palomar Transient Factory (PTF) and Catalina Sky Survey (CSS). The NASA Extragalactic Database (NED) has pioneered integration of data from myriad sources and its holdings are also included within NAVO. The NAVO team works with other NASA and US data providers to ensure that our approach is driven by the needs of the full US astronomy community. In particular NAVO works closely with the United States Virtual Observatory Alliance (USVAO) to ensure that when we assess standards we address not only NASA's needs but the requirements of the national astronomy community.
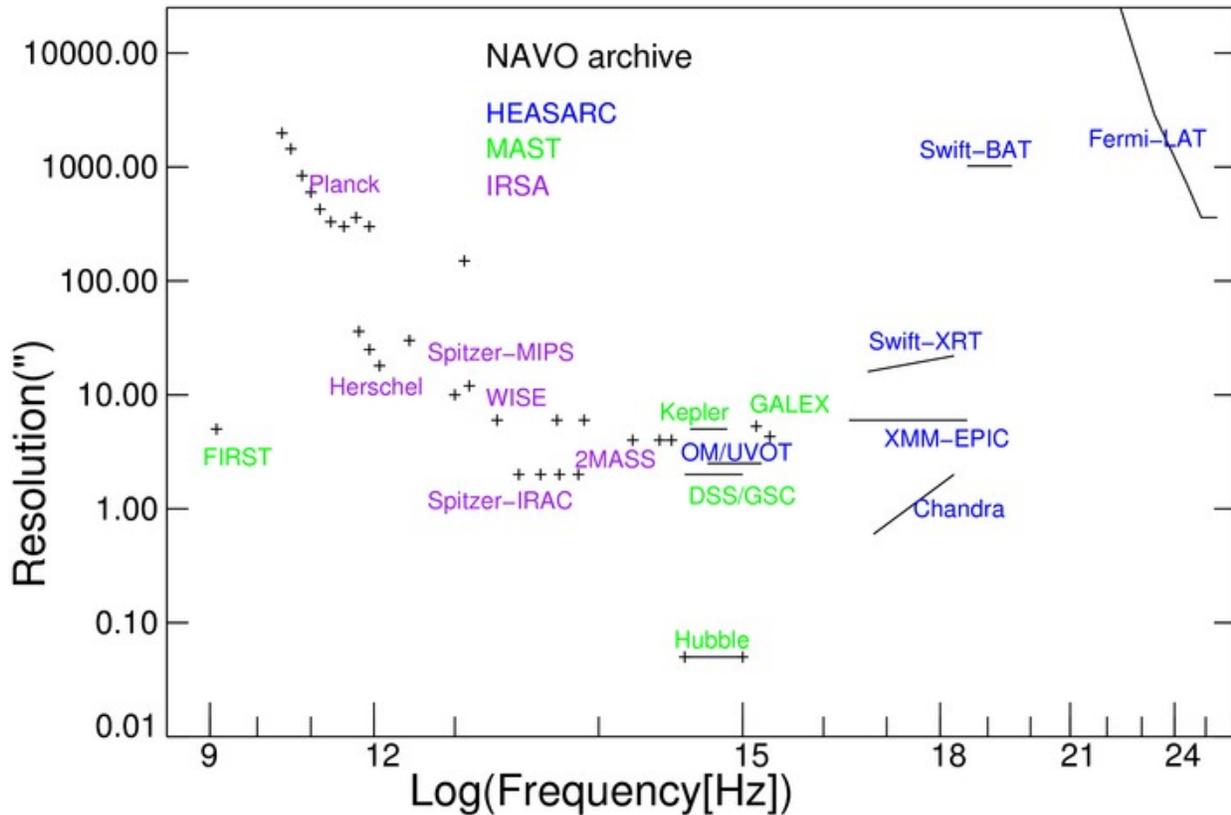

Figure 1. Selected major datasets at NASA archives

NASA archive data is diverse. It spans more than 15 orders of magnitude in frequency. Images' spatial resolutions can range from worse than a degree to better than a tenth of an arc-second. The time resolution may be hours or microseconds. Data can be taken as scans, CCDs or photon events. Files can be kilobytes or gigabytes in size.

Providing a common interface to such heterogeneous data is challenging but enabling astronomers to discover and access our holding through standardized pathways can dramatically enhance the utility of our holdings and will increase the research return on NASA's investment in astronomy. One key challenge for NAVO archives is that the VO interfaces need to work harmoniously with pre-existing interfaces. Adopting VO standards *de novo* is quite different from ensuring that VO interfaces work properly in an existing mature framework.

Nonetheless providing such an interface is essential. We are long past the era where the primary use of our data services is users logging into our systems and manually querying our tables and retrieving our files. The NAVO archives each receive tens of thousands to millions of queries daily the overwhelming majority of which are initiated by external

software clients. Software tools can only interact efficiently and safely with our systems if there is a clearly defined protocol for that interaction. Many such tools are already built using VO standards and using simple query counts we can already see that most of the our queries use these avenues. Many powerful astronomy clients such as TOPCAT [5], Aladin [6] and Iris [7] allow users to query and analyze data immediately using VO protocols. More capabilities are made available on the web or as command line tools each month.

These protocols are also increasingly built implicitly into our custom archive tools. For many years our archive interfaces have provided data in VO compatible formats, but looking just at the HEASARC we find VO interfaces are used to extract information from other NASA sites, to find and query remote catalogs and to allow users to perform correlations with remote data. These VO dependencies are not visible to our users but they take advantage of them every day.

For NASA's holdings to be effectively used in this world where software agents act on behalf of the researcher we need to publish our data using these community standards. Note that in this paper we discuss almost exclusively the requirements on data providers to publish data with only limited reference to client tools like those mentioned above.

2. THE ARCHIVE MODEL

This section discusses our overall archive model, and then examines specific issues that we had to address as we have begun implementing the model. From the perspective of building VO services we divide the data holdings of the archives into three classes (recognizing that the boundaries between them are somewhat porous and ill-defined):

1. *Metadata* describe our datasets and support the data process where users determine which datasets are of interest. These may be stored in a table or more complex hierarchy. They are likely to be queried using a keyword search or by specifying specific position, energy, temporal, or resolution constraints. In this context we are using metadata as what describes missions, instruments and the like not individual observations which are typically indexed using table data.

2. *Archive data* are datasets retrieved as files – including dynamically generated ones. This includes *Observation Data* that describe some aspect of the cosmos and can be used independently of the instrument used to take them. Often these will correspond to higher level (level 2 and 3) data, but can include data like raw CCD frames where positional information has been added to a header. Examples include images, spectra, and time series. Note that the formats of these data may still be instrument specific and they may require special calibrations. Archive data also typically includes *Instrumental data.* These are data whose primary utility other is simply supporting the analysis of specific observation data. These include instrument calibrations, flat fields, housekeeping and orbit data and such.

3. *Table data* are lists of objects, observations, proposals. These are typically stored in relational databases and accessible through query interfaces where users retrieve rows that meet specified criteria. For major surveys tables may be the primary science output of the survey.

The virtual observatory provides different protocols for accessing these kinds of data. Metadata – at the dataset level – is accessed through the virtual observatory **Registry** [8]. Conceptually there is a single VO Registry which may be instantiated at multiple sites. The instances are defined by the interface not the underlying software so they are not necessarily mirrors. The different instances are synchronized so that users may query any Registry instance and get the same result. NAVO supports an instance of the Registry hosted at MAST. Users can query the Registry using either keyword queries or through SQL.

Archive data is currently primarily accessible through image and spectra data protocols**, Simple Image Access (SIA)** [9] and **Simple Spectral Access (SSA)** [10]. These protocols allow a user to query a dataset and find observations that match simple input criteria in position, time and bandpass. Despite the names these protocols can also be used in principle to support datatypes like time series but we have not yet done so.

Tabular data is accessed through two distinct protocols. Simple positional queries – one query per position – can be made through the **Simple Cone Search (SCS)** [11] protocols. The **Table Access Protocol** (**TAP**) [12] is essentially an enhanced SQL interface so that queries can range from a simple search of an individual table to a complex join in some sophisticated schema.

Until recently the VO had few ways to describe instrumental data associated with an observation. However the recently adopted **DataLink** [13] standard allows the user to define any number of files or datasets associated with a specific observation. While there is a suggested vocabulary to be used in defining these relationships, users are free to extend that as appropriate for a given mission. DataLinks can be added to the data returned in most of the protocols above and thus enable the table protocols to be used to index archive data.

Using these building blocks the NAVO team has defined a standard model for publishing a mission in the virtual observatory. Figure 2 illustrates the approach. The NAVO archives have agreed to implement this standard model for all major NASA holdings. While Figure 2 summarizes how the main VO interfaces interact, it hides many of the details that we have had to resolve to standardize the interface. In the next few sections we discuss issues that we have had to address in defining and implementing our interface at the NAVO archives.

Figure 2. Overview of the Standard Model
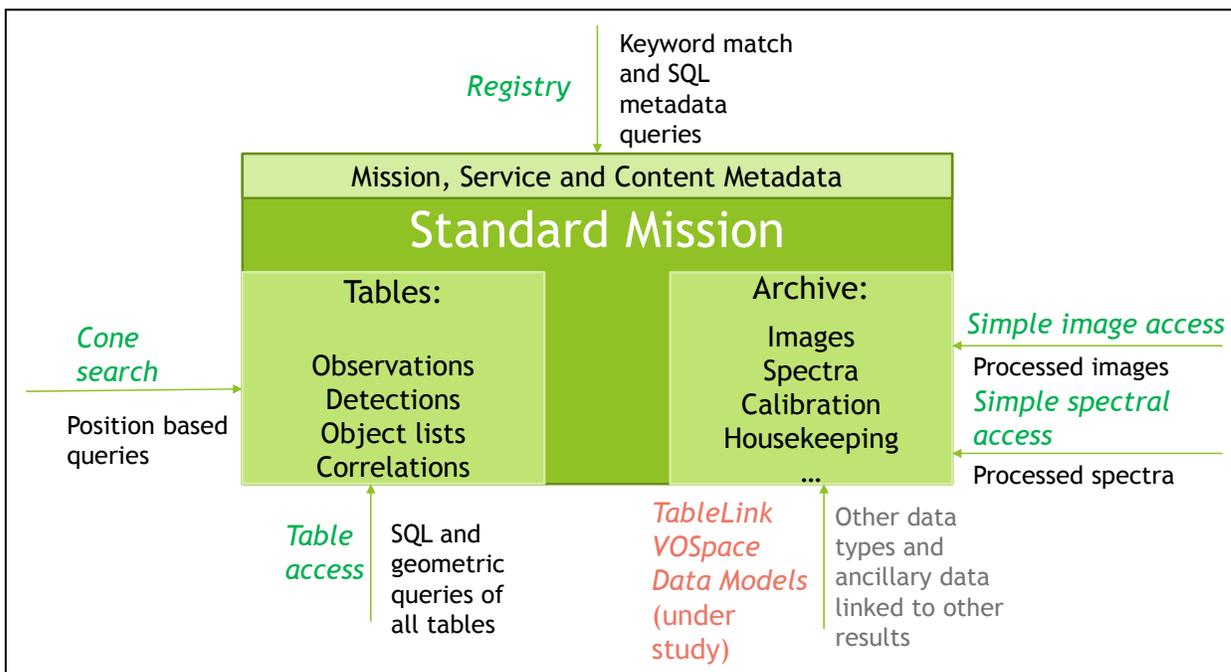
2.1 Issues in registry and metadata access

Publishing data in the registry

While the VO protocols provide a clear mechanism for users querying data after it has been added to the Registry, the mechanism by which data gets into the Registry is left more open. Since the NAVO archive at MAST operates a Registry instance and can manipulate the underlying database directly, adding MAST resources is straightforward. This instance also provides form interfaces that enable users to describe a few resources but become quite cumbersome when used for more than a dozen or so capabilities. The HEASARC and IRSA both support many hundreds of VO resources and concluded that the only feasible way to populate the VO Registry was to build what is called a **Publishing Registry** in somewhat confusing VO parlance. This is not a full instance of the VO Registry, but a tool that is intended to be queried by the full instances to get the resources that a given institution wishes to publish. Although in principle a publishing registry can be queried by users, it is intended only to be seen by the full VO Registry instances. To inform these instances that there is a new location to be searched, an entry must also be made in the Registry of Registries which is found from a link in the IVOA web site. In principle this only needs to a URL to point the new publishing registry. In practice, the Registry of Registries tests the new publishing registry for compliance to basic elements of the Open

Archives Initiative (OAI) interfaces that is used to query the publishing registry and to ensure that the data in the publishing registry matches the required schema. Setting up a publishing registry or full Registry instance is likely to be essential for any major archive.

Supporting complex queries in the registry

The VO's initial protocols for querying the registry enabled only a small number of specific parameter matches, e.g., for the wavelength regime or the publisher. The MAST instance of the registry also supported arbitrary keyword searches. However searches comparable to the kinds of SQL queries users see in other databases were not supported. In April 2016, the NAVO registry added support for the new VO protocols which allow SQL queries of the registry. The NAVO registry now supports three kinds of queries: matches against specified parameters, general keyword searches and full SQL queries.

2.2 Issues in archive access

Selecting protocol versions

Many VO protocols are available in several versions. Typically use of the latest version is appropriate but in some cases the latest version is incompatible with existing implementations. The primary concern currently is the choice of the image access protocol to use. Many NAVO resources are available through version 1 of the protocol, but version 2 allows users to discover data more effectively. Our implementation plan is to finish providing services using version 1 of the protocol and then in coming years transition all services to supporting both versions of the protocol. NED is developing a version 2 implementation of this protocol this year replacing their existing version 1 interface. The experience gained there will be shared with the other NAVO archives.

Fully defining VO protocols

The relationships of VO protocols are complex. As noted above, we plan to transition all of our image access capabilities to use the version 2.0 of the IVOA Simple Image Access protocol. However contrary to its name the dependencies of this protocol are complex. It depends on specific versions of the VO data format, metadata and registry standards and these in turn depend upon other standards. Figure 3 is an analysis of the protocol dependencies of SIA v2. Starting with the primary protocol document in red, each oval represents a document referred to by the standard either directly or through one or more intermediaries. Feedback from this initial review by the NED team was incorporated into the final revision of the document and helped resolve some ambiguities and contradictions. Still understanding how to implement this standard involves a complex skein of interrelated standards. Understanding this fully, and making sure that we have defined exactly what we plan to implement, is a significant effort but necessary to ensure consistency across the NAVO archives.

Implementation of VO Data models

Most recent VO data standards explicitly or implicitly depend upon adoption of a standard observation data model. This **ObsCore** model essentially defines the bounds and resolution of an observation or dataset in spatial, temporal and energy axes. The goal is to ensure that data can be discovered using standardized search methods. In practice this is one of the areas where it is much easier to define appropriate metadata for new missions and observations than to retrofit existing resources into this new framework.

The VO data model does not always easily accommodate data from our missions. It works best with discrete observations with well-defined spatial, temporal and energy boundaries. All-sky monitors, data with fuzzy boundaries, and continuous surveys do not always fit in comfortably. Such datasets may need to be broken up into 'observations' in more or less arbitrary ways. This is not new with the VO. The tiles into which the IRAS's continuous scans were broken into for the IRAS Sky Survey Atlas are an historical example. But it means that incorporation of non-traditional datasets like the Fermi photon database requires some care.

Supporting Instrumental Data

The initial drivers for the Virtual Observatory were strongly biased towards multi-mission analysis where the primary data of interest was high level data with the instrumental signatures largely excised. The VO image and spectral access protocols catered specifically to these use cases. However to get the most out of many mission datasets it is crucial to provide access to detailed supporting instrumental data. With the promotion of DataLink as a standard the NAVO

archives can now begin exploring providing access to the full complexity of space datasets to VO users. Over the next several years as we begin to understand this interface VO interfaces will begin to access the full archive datasets available at NAVO archives. One key issue is the development of clients that that understand and utilize the links.

Figure 3. Dependencies of Simple Image Access V2.0 protocol
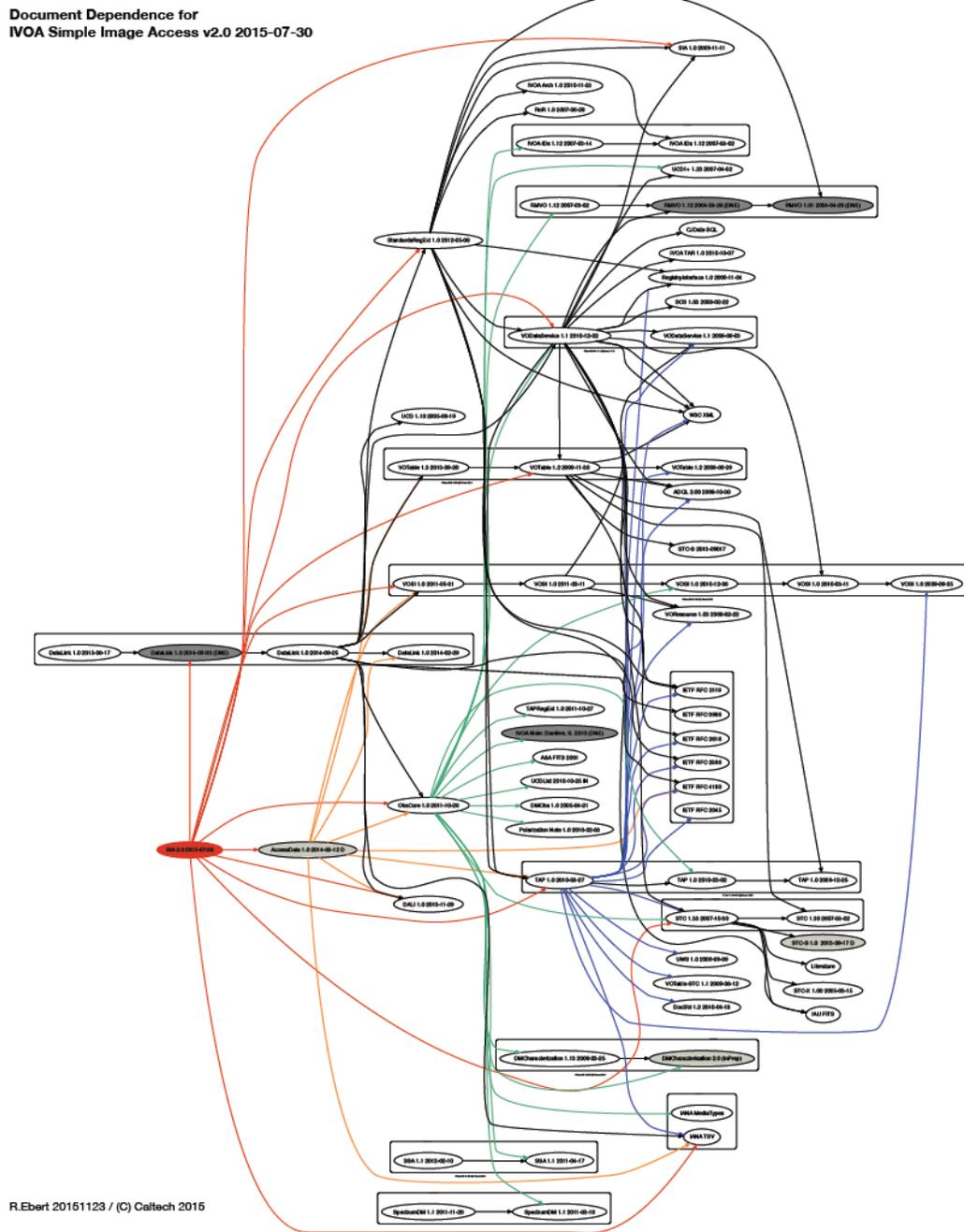
2.3 Issues in table access

Archaic Protocols

As one of the oldest VO protocols, the Simple Cone Search mandates use of obsolete versions of some of the standards that define the meaning of columns. While there has been some sentiment that the cone search standard should be implicitly updated, a consensus has emerged that the cone search standard should be followed strictly. Thus the result returned by a query using the cone search protocol will be slightly different from that using the TAP protocol even when they return exactly the same data. In principle cone searches can be entirely supplanted by TAP queries, but we have chosen to continue to support this protocols since building clients is easier than for TAP.

TAP organization

The VO Table Access Protocol enables clients to do queries of one or more tables. NAVO archives have large numbers of tables and one choice that had to be made was how to publish these data in TAP. Should separate services be provided for tables or groups of tables or should we provide all tables at a given site within a single TAP interface?

The primary advantage for separating TAP services is to enable the metadata descriptions of the TAP service to properly index its content e.g., we could have separate HST and Kepler TAP services, or even a separate TAP service for each HST instrument, where the Registry descriptions can then include quite specific mission information. The advantage of combined services is enabling correlations among the tables. Users can look for regions of the sky seen in multiple instruments or do correlations among object catalogs or combine the datasets in many other ways. Given the scientific utility of these correlations we have chosen to create a single TAP service for each of the NAVO archives.

In principle then, we might have TAP services where we need to describe hundreds of constituent tables. Discovering appropriate tables when a service provides hundreds or thousands of choices would be quite difficult. The IVOA has developed a solution where the metadata for each table is stored separately from the service that queries it. Users find the table of interest and are linked to the appropriate TAP service. Each table is described independently as VO resource in the VO Registry. If the table has position data it will often be associated with a unique VO cone search service. Regardless the table is linked to a single TAP service that spans the NAVO site. In practice most of these tables were already indexed in our registry and we have simply added a link to these entries pointing to the NAVO TAP service.

Handling complex geometries

Another area where TAP allows substantial discretion to the implementer is in the support for special geometry constraints. Since so many queries in astronomy are of the form "find data that is in this region", or "where region A matches region B", the query language that TAP services use adds special geometry functions to standard SQL. In principle these functions can describe arbitrarily complex regions and can be used interchangeably with any other SQL function. In practice doing so leads either to extraordinarily inefficient queries or to ones which are very difficult to translate into the backend server. The NAVO archives have agreed upon a minimum level of support for the geometry functions and are working closely with the IVOA generally to develop agreed idioms for queries which ensure that the spatial indexing is effectively used.

The primary limits to the queries are that the special geometry constraints are easily separable as distinct elements in the query and that functions which in principle support arbitrary regions are restricted to basic circles and polygons (with great circle arcs). This meets all of our common query use cases.

Support for table uploads

TAP allows for optional table uploads where a user uploads a table that can then be used in further queries. An obvious example would be to upload a target list and then search a list of observations or objects for matches. NAVO archives will all support such uploads.

Handling of large tables

For many of our most productive survey missions, e.g., WISE or GALEX, the primary output of the mission is a catalog of the objects detected. There can be hundreds of millions or billions of entries in such catalogs. One of the research goals in NAVO is understanding how we can enable users to access these vast resources. Missions like Chandra and HST have also produced very large object catalogs from heterogeneous observations. All of these have proven to be of

extraordinary scientific value. The NAVO archives are studying how to index these resources to enable the most effective querying and correlation capabilities. This study is underway and is intended to produce recommendations for our common framework for ensuring efficient access to very large tables.

2.4 Future plans

While the plan for a common VO interface to the NAVO archives was included in the proposal for NAVO, the details of this interface have been developed as part of the project and are subject to annual review. In our first year the basic definition of the elements of the interface were defined, the gaps between our goals and current implementation were assessed, and a schedule for filling these gaps was determined. In the second year, a clearer understanding of the interrelationships of the VO protocols was achieved and services targeted specifically to fill in the missing interfaces began to come online. In the next year most of the remaining holes in basic coverage will be addressed. However we will also being transitioning to more advanced VO protocols. Three protocols that we have not yet fully incorporated into the common model should be of particular interest as we re-assess the common interface this year: **DataLink**, **ObsCore** and **VOSpace**.

The DataLink protocol allows us to associate a row of a table returned in any protocol with a set of one or more data products. These products can be organized in a hierarchy. This allows VO access to mirror the non-VO standards where data tables index the archive content and full observation datasets, not just simple images or spectra, are returned.

ObsCore[14] is the core of the virtual observatory data model and will be gradually incorporated in the system to support the newest version of the VO data access.

VOSpace[15] addresses a new aspect of users' interactions with the archive. As data volumes in archives and tables get ever larger, it behooves the archives to enable users to store intermediate results close to the data rather than requiring downloads and the re-uploading data at every step. VOSpace is the IVOA approach to cloud storage. Over the next year we are exploring how VOSpace may be used and whether alternatives such as WebDAV might be easier to implement and more effective for our users.

Until the present VO protocols have focused heavily upon the discovery of resources: dataset collections can be found in the registry, the image and spectral access protocols return matching observations. But the VO has provided relatively limited assistance in the making data access easier. If an image service returns a huge image, there is no standardized way to extract the small subset a user might be interested in. Protocols that would enable the kinds of subsetting and filtering that could make it much easier for users to access very large data are now under development in the IVOA. This effort is being carefully monitored by the NAVO team.

3. PRIMARY NASA HELD DATA SETS

Many of the elements of our standard interface have been already be implemented or will shortly be available. Table 1 shows the status of protocol implementation of almost 60 major missions and NASA-held data sets. Datasets where the protocol is available in June 1, 2016 are indicated by a check, ✓. If a mission cannot sensibly support a protocol (e.g., spectral access to an imaging mission), or for some older missions where the effort of retrofitting full VO access may not be cost effective, a protocol may be marked with ✗ which indicates that we do not do not currently plan to implement the protocol for that dataset. Otherwise the year when implementation is scheduled is shown. Missions where NASA is the primary sponsor are shown in **bold**. Many smaller missions and datasets have been omitted.

We give a common name – usually an acronym – for each mission and the general spectral regimes covered. A single mission line, e.g., HST or Spitzer, may incorporate many different datasets. The subsequent columns indicate respectively: the status of the registration of metadata; the ability to do simple positional queries of observation and object catalogs associated with the mission; whether image data is available through the VO; whether spectral data is available and whether the mission tables are queryable through the standard TAP (i.e., enhanced SQL) interface. Metadata and simple positional queries are available for all datasets, image access and general SQL query capabilities will be almost universal by the end of this year (excluding non-imaging datasets of course). While some missions already have spectral access much more will be available by the end of next year. Our plan has been to basically finish our first level of comprehensive access in the first three years of NAVO.

Mission	Regime	Registry	Cone	Image	Spectra	Table
2MASS	IR	✓	✓	✓	✗	✓
AGILE	X-ray	✓	✓	✗	✗	✓
AKARI	IR	✓	✓	✗	✗	✓
Ariel V	X-ray	✓	✓	✗	✗	✓
ASCA	X-ray	✓	✓	2016	2017	✓
BBXRT	X-ray	✓	✓	✗	2017	✓
BEFS	UV/EUV	✓	✓	✗	✓	2016
BeppoSAX	X-ray	✓	✓	✗	✗	✓
BLAST	Sub-mm	✓	✓	✓	✗	✓
BOLOCAM	Mm	✓	✓	✓	✗	✓
Chandra	X-ray	✓	✓	✓	2017	✓
COBE	CMB/IR	✓	✓	✓	✗	✓
Compton	Gamma-ray	✓	✓	2016	2017	y
Copernicus	UV/X-ray	✓	✓	✓	✗	2016
COS-B	Gamma-ray	✓	✓	✗	✗	✓
COSMOS	All	✓	✓	✓	✗	✓
DENIS	IR	✓	✓	✗	✗	✓
DSS/GSC	Optical/IR	✓	✓	✓	✗	2016
Einstein	X-ray	✓	✓	2016	2018	✓
EUVE	EUV	✓	✓	✓	✗	✓
EXOSAT	X-ray	✓	✓	2017	2018	✓
Fermi	Gamma-ray	✓	✓	2016	2017	✓
FUSE	UV	✓	✓	✗	✓	2016
GALEX	UV	✓	✓	✓	✗	2016
GINGA	X-ray	✓	✓	✗	✗	✓
HEAO-1	X-ray	✓	✓	✗	✗	✓
Herschel	IR	✓	✓	✓	✗	✓
HETE-2	Gamma-ray	✓	✓	✗	✗	✓
HPOL	Optical/IR	✓	✓	✗	✓	✓
HST	IR/Optical/UV	✓	✓	✓	✓	2016
HUT	UV	✓	✓	✗	✓	2016
IMAPS	UV	✓	✓	✓	✗	2016
INTEGRAL	Gamma-ray	✓	✓	2017	2017	✓
IRAS	IR	✓	✓	✓	✗	✓
IRTS	IR	✓	✓	✓	2017	✓
ISO	IR	✓	✓	✓	✗	✓
IUE	UV	✓	✓	✗	✓	2016
Kepler	Optical	✓	✓	✓	✗	2016
MSX	IR	✓	✓	✓	✗	✓
NuSTAR	X-ray	✓	✓	✓	2017	✓
Planck	IR	✓	✓	✓	✗	✓
PPMXL	IR/Optical	✓	✓	✗	✗	✓
PTF	Optical	✓	✓	✓	✗	✓
ROSAT	X-ray/EUV	✓	✓	✓	2017	✓
RXTE	X-ray	✓	✓	✗	2017	✓
SAS-2	Gamma-ray	✓	✓	✗	✗	✓
Spitzer	IR	✓	✓	✓	2017	✓
Suzaku	X-ray	✓	✓	2016	2017	y
Swift	Opt/UV/X/g-ray	✓	✓	✓	✗	✓
TUES	UV	✓	✓	✗	✓	2016
UIT	UV	✓	✓	✓	✗	2016
USNO	Optical	✓	✓	✗	✗	✓
Vela 5B	Gamma-ray	✓	✓	✗	✗	✓
VLA-FIRST	Radio	✓	✓	✓	✗	2016
WISE	IR	✓	✓	✓	✗	✓
WMAP	CMB/IR	✓	✓	✓	✗	✓
WUPPE	UV	✓	✓	✗	✓	2016
XMM-Newton	X-ray/UV/Optical	✓	✓	✓	2017	✓
NED	All	✓	✓	✓/2016	2019	2018

Table 1. Implementation of VO protocols for NAVO datasets.

4. STATUS AND PLANS

Since its inception NAVO has collected statistics on the usage of VO protocols at each of the NAVO sites. Even in 2014, VO requests represented about half of the queries presented to the HEASARC archive and played a significant role at MAST. During 2016 VO usage at IRSA has increased dramatically so that currently most queries there are being made through VO methods. VO access at NED still represents a small fraction of total queries, but NED high level object queries are not supported by VO protocols. Where there is a direct comparison, e.g., between NED SED services which are available through both VO and non-VO interfaces, the VO queries are comparable.

It is hard to quantify the science relevance of these numbers. Small numbers of users making large numbers of queries dominate both the VO and non-VO query load. Nonetheless the general trend seems to be that VO services will be a major and increasing fraction of our query load.

With the increasingly wide adoption of TAP and the prospect of general implementation of DataLinks the VO protocols are now powerful enough to enable full distribution of NASA data through VO protocols. In our first 20 months the NAVO archives have defined and substantially implemented standardized data access to the major NASA astronomy holdings. In the coming years we hope to build upon this foundation using DataLinks and the VO data models to not only allow consistent and comprehensive access to our datasets but to simultaneously support access to the full mission holdings for each mission, and to enable completely mission independent queries of the NASA astronomy holdings.

REFERENCES

[1] Brunner, R.J., Djorgovski, S.G., Szalay, A.S., [Virtual Observatories of the Future], ASP Conference Proceedings, Vol 225, Astronomical Society of the Pacific, San Francisco (2001)

[2] Quinn, P.J., Gorski, K.M., [Toward an International Virtual Observatory], ESO Astrophysics Symposia, Springer-Verlag, Berlin/Heidelberg (2004)

[3] Hanisch, R., Quinn, P, "The International Virtual Observatory", International Virtual Observatory Alliance 2003, http://www.ivoa.net/about/TheIVOA.pdf (22 May 2016)

[4] Graham, M.J., Fitzpatrick, M.J., McGlynn, T.A., [The National Virtual Observatory: Tools and Techniques for Astronomical Research], Astronomical Society of the Pacific, San Francisco (2007)

[5] Taylor, M.B., "Visualizing Large Datasets in TOPCAT v4", ADASS XXIII, ASP Conference Series 485, 257 (2014).

[6] Boch, T., Oberto, A., Fernique, P., Bonnarel, F., "Aladin: An Open Source All-Sky Browser", ADASS XX, ASP Conference Proceedings 442, 683 (2011)

[7] Laurino, O., et al., "Iris: An extensible application for building and analyzing spectral energy distributions", A&C 7, 81 (2014)

[8] Demleitner, M, Greene, G., Le Sidaner, P., Plante, R.L., "The virtual observatory registry", A&C 7, 101 (2014)

[9] Dowler, P, Bonnarel, F, Tody, D., "IVOA Simple Image Access Version 2.0", IVOA Recommendation 23 December 2015 (2015)

[10] Tody, D., et al., "Simple Spectral Access Protocol Version 1.1", IVOA Recommendation 10 February 2012 (2012)

[11] Plante, R., Williams, R., Hanisch, R., Szalay, A., "Simple Cone Search Version 1.03", IVOA Recommendation 22 February 2008 (2008)

[12] Dowler, P., Rixon, G., Tody, D., "Table Access Protocol Version 1.0", IVOA Recommendation 27 March 2010 (2010)

[13] Dowler, P., Bonnarel, F., Michel, L., Demleitner, M., "IVOA DataLink Version 1.0", IVOA Recommendation 17 June 2015 (2015)

[14] Tody, D., et al., "Observatory Data Model Core Components, its Implementation in the Table Access Protocol Version 1.0", IVOA Recommendation 28 October 2011 (2011)

[15] Graham, M., et al., "VOSpace specification Version 2.0", IVOA Recommendation 29 March 2013 (2013)