



International

Virtual

Observatory

Alliance

The IVOA in 2005: Assessment and Future Roadmap

IVOA Technical Coordination Committee

Version:

2005 July 13

Author:

Roy Williams

In consultation with:

IVOA Technical Coordination Committee:

Francoise Genova, Bob Hanisch, Gerard Lemson, Tony Linde,
Jonathan McDowell, Tom McGlynn, Reagan Moore, Maria Nieto-Santisteban, Francois
Ochsenbein, Masatoshi Ohishi, Andrea Preite Martinez, Guy Rixon, Yuji Shirasaki,
Doug Tody

Abstract:

This document is the result of a study by the IVOA Technical Coordination Committee (chair R. Williams) with the intention of coordinating the IVOA Working Groups and Interest Groups. The study was commissioned by the Executive Committee in May 2005 with the objectives of:

- Building a roadmap for the IVOA that is a union of roadmaps for the Working Groups and Interest Groups.
- Ensuring productive crosstalk of the WG/IG so that workpackages cover relevant ground, but also do not overlap.
- Evaluating dependencies of one WG/IG on another and minimizing impact.
- Attaching milestones to the WG/IG roadmaps, representing planned achievements and target dates.
- Ensuring an effective evaluation of proposed standards during the RFC period.
- Providing a continuous reporting checkpoint to the IVOA Executive Committee on roadmap status.

Initial Overview Roadmaps

The current roadmap situation (July 05) is summarized in Table 1, the Working Groups and Interest Groups, and Table 2, the proposed roadmap for each WG/IG. Since one of the main objectives of the IVOA is production of standards documents, the status of these documents is called out in terms of what type of document is being produced and the stage it has reached in that production.

In Table 2, documents that are in progress or in the future are denoted thus:

- **Note:** No approval process required.
- **WD:** Working draft.
- **PR:** Proposed Recommendation
- **REC:** IVOA Recommendation

The status of a document can be:

- **NYS:** Not Yet Started
- **InWG:** Preparation within WG, meaning that a draft is being circulated among a subset (or all) of the WG, and that action is on the WG chair to ensure progress.
- **RFC:** Submitted for comment outside WG, meaning that the document is in the official "Request for Comment" from all IVOA members, and that the request period has a given expiry date.

In this scheme, a document flows through the scheme as follows:

NYS – WD inWG – WD in RFC – PR in RFC -- REC

Working Group Chair Responsibilities

- Each WG must have a clear Roadmap in a standard form - with planned achievements versus target dates (i.e. milestones)
- WGs should pay close attention to the top-level Technical Milestones, making sure each relevant milestone is inside the WG roadmap.
- There should be a checkpoint at each Exec Meeting and at each Interop Meeting
- For each checkpoint, the WG chair should provide (i) a very short text report (1-2 paras) (ii) a progress statement on each element of their roadmap
- The above reports will be requested 2 weeks in advance from the IVOA.

In addition to the above responsibilities for her own Working Group, the Chair is also responsible for active comment (1-3 paragraphs) on each request for comment (RFC) that has been issued by another Working Group.

Interest Group Chair Responsibilities

- Reporting by IGs should be relatively low key and informal. This informality is a key distinction between WGs and Igs. WGs are much more work, and need to deliver a product.
- IGs should provide verbal reports at each Interop meeting.
- The Interop organising committee should request these several weeks before the Interop Meeting.

General Recommendations

One of chief reasons for success of the IVOA and its member organizations has been the readiness to make fast prototypes and pragmatic tools that have immediate value to astronomers. But in the longer term, there is a need for coherent, extensible, flexible architecture – the strategic approach. This issue emerges in many of the working groups: the DAL is a pragmatic approach (conesearch and its children), whereas Data Models is a strategic approach. In the former, it is difficult to extend the tool that was built with limited goals, but in the latter, it is difficult to get the completed data model to be used in pragmatically-built prototypes. The semantics work is similarly divided into UCD (pragmatic) and ontology (strategic). The greatest challenge of the IVOA is to approve user-driven, pragmatic efforts, and at the same time build a comprehensive architecture. **Recommendation:** The way to achieve this is through *layering*: each data model or protocol should be split into layers, so that the simplest layer is obviously useful, with a progression of sophistication until the entire architecture is exposed. Thus services should have *levels of compliance*, data models and schema should have *levels of sophistication*.

In terms of overlap of workpackages, there seem to be multiple data models and representations that express the metadata of a catalog/table. VOTable is a mechanism for expressing models in an ad hoc, easily transportable way, as well as providing a data modeling language. The Registry group is working with the Data Models group to make another format for expressing table/catalog, a very simple way of recording table and column details (name of each and the UCD of each column). When the catalog/table data model appears, it will be incorporated into the registry description for those types of resource and the current mechanism will be deprecated. The same can be said of any other type of resource: as soon as there is a workable data model, the resource schema will use it.. Each of these groups must explain why they are not using what the other two groups have done. **Recommendation:** a cross-WG effort should be initiated to integrate approaches to tables and catalogs.

As XML becomes the lingua franca of the IVOA, there are obviously a lot of XML schema to validate documents. These schema are currently located in various URLs and web sites, with ivoa.net/xml as the preferred place, and the IVOA document coordinator as manager. Once there are many documents that point to a given schema location, it becomes impossible to change, so we should be quite sure this is being done correctly. **Recommendation:** An effort should be initiated to decide on schema locations, how that space is managed and who can add/modify/delete in that area.

In the area of security and authentication, there is an urgent need for protocols, prototypes, and trust servers. New data streams coming online should be using IVOA protocols, but sequestration is required and the data will not be public for a year. Expensive resources can be used for IVOA services, but only if the owners can know who is using the resources. Distributed storage is viable only if customers can be sure that the data is safe from mistakes and hackers.

A significant thrust for the future of the VO Registry must be the human and workflow side, how records are created, checked automatically, checked by human, cleaned syntactically and semantically, monitored and maintained, how the owner of records can change them. While this is not perhaps an IVOA activity, we would be happy to see the Curation and Preservation IG work with the Registry WG to build a document of best practices.

Specific Working Groups

The **Applications IG** is a forum for discussion of applications and portals that rely on IVOA services and protocols.

The **Data Access Layer** group has focused on extending the original conesearch into useful, simple, deployable services that have seen great take-up in the community. The latest of these, "simple spectral access interface", is used for spectra, time series, and SEDs and is based on common spectrophotometric data model. As the DAL services gain sophistication, there should be plans for backward compatibility. Further, as with other WG, the standards should have "levels of compliance", perhaps in terms of core capability and optional extensions. This is to ensure that it is still a simple matter to create a simple service.

The DAL group is working closely with the Data Models group to allow rich metadata to be attached to complex datasets from multiple instruments (SED, characterization DM). It would be good for the IVOA to work more closely with the major suppliers of such data, such as NASA's NED system. The DAL group is also working closely with the Query Language group so that specific DAL queries can be translated to the more general ADQL. The DAL group is waiting for the Grid and Web Services group for distributed storage, security, and asynchronous service protocols and implementations.

The **Data Curation and Preservation IG** is a forum for evaluation of metadata formats and methods, ways to integrate IVOA with digital libraries, and evaluating preservation environments such as Dspace and Fedora.

In the **Data Models** group there seems to a tendency to produce documents without a plan to integrate into the architecture. Part of the problem is that the models are very comprehensive and emphasize the most sophisticated cases over the most ordinary. It is then difficult to make software that implements the data model -- because it is so comprehensive. With no software, there is then no take-up by the community. Perhaps if a data model were simpler, then it would be much easier to produce the implementing software. Another tendency is to make Data Models ab initio, rather than as a rationalization of what is already being done, or as a response to requirements. This can lead to abstract discussion and complex models that may be only tangentially relevant to working astronomers.

Another effect of a sophisticated data model is the impression in the community that all levels of complexity must be understood before any part of it can be used. It would be better to have data models that can be used at different levels of sophistication. In the case of the **Space-Time Coordinate** system specification, now ready to become a Proposed Recommendation, there is a general unease: while it is agreed that it is a jewel of the IVOA, it can still fail to gain acceptance without simple, practical paths to use it. Many astronomers feel that a coordinate system name (eg "J2000", "Galactic"), together with two numbers, is enough to express a position in the sky. They understand the meaning and the deficiencies of that simplicity. But STC must be layered or sequenced in sophistication, so that the simple position above can be expressed simply with a simple schema; so that pieces of STC can be used independently; and of course that a "sophisticated position" can be expressed fully.

The **Event** group has produced a semantic specification that is in WD, how to represent an observation of an immediate astronomical event with a view to follow-up. The specification uses Space-Time Coordinates for position, and expects to have Event servers listed as Registry resources.

The **Grid and Web Services** WG is responsible for four critical path items:

- Security and trust protocols will be vital for any fully operational Virtual Observatory, and progress is underway with prototypes in development.
- Interfaces to asynchronous, computationally intensive services, (Universal Worker Service specification, based on WSRF), and
- Distributed storage (VOStore). This is another another foundation stone for deploying compute-intensive services that can be connected in a workflow. VOStore should be carefully defined so that data store can be indirect, not controlled by VOStore, yet the metadata still well synchronized with the data store.
- Asking a service about itself (support interface). This includes how services should be written, for how usage can be logged, how services can be automatically checked for health. While it is desirable for all IVOA services to comply to this, it would not be desirable to reject services that do not comply.

There seems to be a lot of dependency on the activities of GWS, and the WG should perhaps be split. The advantages would be the hope of greater participation, and also the efficiency of a smaller group.

The **Query Language** group (VOQL) is creating ADQL (Astronomical Data Query Language) as an extension of the relational model with sky regions and cross-match capabilities. It is using the Space-Time Coordinate specification (STC) to define such regions, and to address databases in the distributed storage system (VOStore).

Another major thrust of the group is the SkyNode protocol, the next generation of the DAL Cone Search protocol, providing federated access to distributed astronomical databases. SkyNode is written as a set of SOAP services, therefore there is a strong relationship with the GWS group regarding Support Interfaces, VOStore, Async interfaces and Security protocols. Large scale cross-match will require implementation of VOStores and Async protocols.

The VOQL group is also collaborating with the Data Access Layer group to build the pragmatic DAL query using the more general ADQL. In this way the VO user can move smoothly from something as simple as a cone-search, to something as sophisticated as distributed, cross-matched catalogs using ADQL.

The **Registry** group has been very successful in creating the distributed VO registry infrastructure. The resource data model is sophisticated, and getting more so, and there is a critical need to understand how resources are ingested into the registry, then checked and corrected on a regular, automatic basis. The semantic nature of the metadata has converged to standard, and now the syntax and transport is also being standardized.

The VO Registry structure is proving robust and flexible. New types of resource are appearing, for example the workflow components from the UK VO, or the event aggregators of the real-time VOEvent group. Each new resource type requires distributed human effort and could be streamlined. Similarly, the harvesting from registry to registry should be streamlined. The recommendation is to build a "Registry of Registries" to hold the linkage information.

Another area for the Registry WG to discuss and develop is the uses to which the registries and their contents are to be put. Surveying how registries are searched and how resource metadata are used in applications, to incorporate ways of ensuring such usage is efficient and accurate (eg introduction of simple way to get single resource record in latest version of RI doc). In this, a long-term developing area will be semantic resource discovery.

The **Semantics IG** is a discussion forum for the idea of introducing semantics into astronomy, including links with data models, resource discovery, query languages, intelligent workflow, etc. Individual aspects may then be floated off into specific WGs.

The IVOA has a formal system for creating consensus documents, overseen by the **Standards and Documents** group. This process relies on IVOA members taking the initiative to understand that a new document is in the RFC stage, to read that document, and engage in the comment process. It would be better if that initiative were taken together by the document author and IVOA document coordinator. A real improvement is the recent appearance of the “Document Matrix” at the IVOA web site¹, showing which documents are at which stages in the recommendation pipeline. **Recommendation:** Further improvements in the practical aspects of the process, including a clear path from RFC document to the place where comments can be made. Each WG/IG chair should be required to comment on or approve each RFC document, and the web pages could show who of those have not done so, and send automatic reminders.

The **Table WG** has achieved a great deal with the specification, dissemination, and wide take-up of this important representation over the international community. As noted in the previous section, we recommend an IVOA-wide overview of the modeling of tables and catalogs.

In the **Theory IG**, there are four main activities:

- Large scale/cosmological simulations (Shaw, Lemson);
- Medium scale/galactic simulations (DeYoung et al);
- Theoretical spectra (JHU/HVO, Solano/Osuna);
- Atomic (Dubernet).

The **UCD Working Group** is making human workflow for the updating and enhancement of the UCD words list. This important work will serve as a template for other activities in IVOA that maintain such “approved” lists. The UCD group is also investigating the creation of a vocabulary/ontology covering astronomical objects, and events and phenomena associated with those objects.

¹ <http://www.ivoa.net/Documents/docMap.html>

Table 1: IVOA Working Groups and Interest Groups

Working/Int. Group	Chair	Current priorities
Applications	IG Tom McGlynn	Various application news.
Data Access Layer (DAL)	WG Doug Tody	Spectral Energy Distribution (with DM). Simple Spectral Access 0.9WD, Extending Simple Image Access, 3D data and characterization
Data Curation and Preservation (DCP)	IG Francoise Genova, Reagan Moore	Metadata formats and methods. Evaluating Preservation environments (eg Dspace, Fedora). Curation/maintenance of registries?
Data Models (DM)	WG Jonathan McDowell	Spectral Energy Distribution (with DAL) Characterization (of observations) DM Space-Time coordinates (STC). Catalog DM Provenance (of observations) DM Spectral line (atomic line) DM
Event	WG Roy Williams	Event Semantics WD 1.0 and schema. Prototypes and transport.
Grid-Web Services (GWS)	WG Guy Rixon	Security, trust, single sign-on. Prototypes. VOStore and VOspace. Asynchronous services and WSRF. Logging and support for services.
Query Language (VOQL)	WG Maria Nieto Yuji Shirasaki	Astronomical Data Query Language (ADQL) as XML and script. SkyNode Interface methods. Integration with DAL
Registry	WG Tony Linde	Resource Metadata document 1.1, VOResource (& associated) schema 1.0, and Registry Interface specification 1.0 Registry of registries Registering general services and applications. Query languages for the registry.
Semantics	IG Tony Linde	A think-tank -- role of ontology in data models, resource discovery, query languages, intelligent workflow
Standards and Documents (SD)	WG Bob Hanisch	Improved workflow for RFC process
Systems Architecture & Technical Coordination (TCC)	Roy Williams	Technical Coordination Committee: overlap, dependencies, RFC process.
Table	WG Francois Ochsenbein	Parsers, implementations and bug fixes.
Theory	IG Gerard Lemson	Large scale/cosmological (Shaw, Lemson); Medium scale/galactic (DeYoung et al); Theory spectra (JHU/HVO, Solano/Osuna); Atomic (Dubernet).
UCD	WG Andrea Preite-Martinez	Updating and agreeing UCD list. Workflow for changes to list. Services and use cases for UCD.

Table2: **IVOA WG Roadmap July 2005**

Date	WG/IG	Standard	Status	Responsible
May-04	DAL	Simple Image Access-V1.0	PR RFC	Tody, Plante
Jul-05	DAL	Simple Spectral Access-V0.9	WD inWG	Tody, Dolensky
Oct-05	DAL	Simple Linelist Access-v0.1	WD inWG	Dubernet, Osuna
Oct-05	DAL	Simple Spectral Access-V1.0	WD NYS	Tody, Dolensky
Jan-06	DAL	Simple Image Access-V1.1	WD inWG	Tody
	DAL	Time Series (is included in SSA)		
2006	DCP	Preservation Recommendations	Note	Genova, Moore
Apr-05	DM	Data Model for Observation	Note	McDowell et al
Jul-05	DM	Spectrum Line Lists-V0.1	WD inWG	Dubernet, Osuna
Sep-06	DM	Characterisation-V0.9	WD inWG	Bonnarel, Louys
Sep-05	DM	Spectrum 0.9	WD inWG	McDowell
Oct-05	DM	Space Time Coordinates-V1.21	PR RFC	Rots
Dec-05	DM	Spectral Energy Density-V1.0	WD inWG	McDowell, Tody
Dec-05	DM	Catalog-V0.5	WD inWG	Osuna
Jan-06	DM	Characterisation-V1.0	WD inWG	Bonnarel, Louys
2006	DM	VOQuantity-V1.0	Note	McDowell, Berry, Dowler, Thomas
Jul-05	Event	VOEvent -- Immediate Event Reporting-V1.0- Semantics	WD inWG	Seaman, Williams
Dec-06	Event	VOEvent -- Immediate Event Reporting-1.0-Transport	WD NYS	
Sep-05	GWS	VO-Support Interface-V1.0	WD inWG	O'Mullane, Rixon, Thakar
Nov-05	GWS	VO- Web Service Interoperability-V1.0	WD inWG	Schaaf
Dec-05	GWS	VOSTore (Distributed Storage) V1.0	WD inWG	Graham, Morris, Plante
Jan-06	GWS	Single Signon Authentication V1.0	WD inWG	Graham, Plante, Rixon
Jan-06	GWS	Universal Worker Service -V1.0	WD inWG	Rixon
Jan-07	GWS	VOSpace (Federated Distributed Storage) - 1.0	WD NYS	
Oct-03	SD	IVOA Document Standards 1.0	REC	
Jul-05	VOQL	Astronomical Data Query Language -v1.01	WD RFC	Nieto, Shirasaki
Jul-05	VOQL	SkyNode (Queries on joint catalogs) -v1.01	WD RFC	Nieto, Shirasaki
Sep-05	VOQL	Integration of DAL and ADQL concepts		Nieto, Shirasaki
Sep-06	VOQL	Astronomical Data Query Language V2.0	WD NYS	

Date	WG/IG	Standard	Status	Responsible
Mar-05	Registry	VO-Identifiers V1.10	PR RFC	Plante
Jun-05	Registry	Resource Metadata V1.1	PR RFC	Hanisch
Jul-05	Registry	VOResource (schema)-V1.0	WD	Plante
Nov-05	Registry	Registry Of Registries -V1.0	Note	Plante
Feb-06	Registry	Registry Interface-V1.0	WD inWG	Benson
Jun-04	TCC	VO Architecture	Note	Williams
Oct-05	TCC	"What is the VO":	Note	Lawrence, Walton?
Dec-05	TCC	Update to architecture: Compute services	Note	Williams, Rixon
Aug-04	Table	VOTable-V1.1	REC	Ochsenbein
Aug-05	Table	VOTable-V1.2	WD inWG	Ochsenbein
Oct-05	Table	Review status of VOTable working group.		Ochsenbein
Aug-05	Theory	Theory use cases	Note	Lemson
Oct-05	Theory	Theory requirements	WD NYS	Lemson
Oct-04	UCD	Unified Content Descriptors 1+ V1.06	PR RFC	Derriere, Preite Martinez
Jun-05	UCD	Create vocab and tech editorial boards		
Jul-05	UCD	UCD-vocabulary with editorial board		