International

Virtual

Observatory

Alliance

The IVOA in 2006: Assessment and Future Roadmap

IVOA Technical Coordination Committee

Version:
2006 May 10

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 and revised in May 2006, 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 (May 06) 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 labeled by their status in the IVOA document sequence:
- **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
- **WD**: A Working Draft is available on the IVOA Documents page, at level 1,0 or greater.
- **PR**: The chair of the Working Group has notified the Technical Coordination Committee and the IVOA Document Coordinator, and a 4-week comment period has started, with proper instructions for how to comment. This cycle can happen several times.
- **REC**: The Executive Committee of the IVOA has moved this to a Recommendation.

In addition to the above document categories, working groups or other groups can also submit a **Note**, which is not an explicit part of the standards process.

Working Group Chair Responsibilites

- 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

In July 2005, these recommendations covered
- Levels of compliance for services and data models, so that simple things can be done simply at a lower level, yet complex things are possible at a higher level.
- A unified approach to the representation of table and catalog metadata, unified across Registry, Data Model, and VOTable working groups.
- Prototyping and implementation of security and authentication infrastructure.
- The human workflow of a Registry, how Authors (of registry records), Publisher, Curators, and Domain Experts can interact effectively.

The unification of catalog metadata seems to be advancing. Security and authentication is being implemented in several new efforts.

A year later, in May 2006, some IVOA standards have matured, so that a new version is appearing (eg. the Simple Image Access Protocol, the VO Query Language). It is hoped that the older, simpler version can be retained in addition to the new version, because (a) many sites will continue using the original standard, and (b) the older, simpler, protocol can often do everything that is needed. Thus we see versioning as equivalent in many ways to the "levels of compliance" noted above: the new version should not be seen as a replacement, but an enhancement. We **recommend** continued support of older versions in addition to the newer.

For the last of these four points above, we note that the registry work is not being directed toward usability *per se*, but rather to schema evolution. A new plan will be discussed at the May 2006 Interop, that elaborates the idea of *Service* into a family: the parent *Service* contains *Interfaces* and *Capabilities*. We **recommend** that the IVOA Exec should decide on the future of the Registry, and the level of complexity that it can sustain; on whether schema development should be halted to allow long-term implementation; to what extent the Registry has fixed functionality, rather than an evolvable and dynamic system. We also recommend that the registry WG define the scope of the registry in terms of the variety of supported metadata, it sophistication, and its level of granularity.

The VOSpace effort within the Grid/Web services working group is building semantics, schema, interface, and prototype. While it is clear that astronomers using the VO will be interested in storing data in networked resources, the IVOA exec may wish to take the challenge once again of asking why the astronomical community is building this system, and what other systems exist that could be directly utilized. We **recommend** the formation of a *VOSpace Use Cases* document, to more closely define the direction of this fine effort, and to differentiate it from related efforts in the grid community.

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 (VOSpace). 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 atreamlined. 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/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. The charter includes links with data models, resource discovery, query languages, intelligent workflow

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.

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).

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/UCD	WG Andrea Preite-Martinez	Updating and agreeing UCD list. Workflow for changes to list. Role of ontology. Standard vocab for Process/Objects
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).

Table2: **IVOA WG Roadmap May 2006**

Date	WG/IG	Standard	Status	Responsible
May-04	DAL	Simple Image Access-V1.0	PR	Tody, Plante
Oct-05	DAL	Simple Linelist Access-v0.1	inWG	Dubernet, Osuna
May-06	DAL	Simple Spectral Access-V1.0	inWG	Tody, Dolensky
May-06	DAL	Simple Image Access-V1.1	inWG	Tody
Jul-05	DM	Spectrum Line Lists-V0.1	inWG	Dubernet, Osuna
Feb-06	DM	Characterisation-V1.0	WD	Bonnarel, Louys
Mar-06	DM	Space Time Coordinates-V1.3	PR	Rots
May-06	DM	Spectral Energy Density-V1.0	WD	McDowell, Tody
2006	DM	VOQuantity-V1.0	Note	McDowell, Berry, Dowler, Thomas
Jul-05	Event	VOEvent 1.0 – Semantics and Schema	WD	Seaman, Williams
May-06	Event	VOEvent --Transport	Note	Seaman
May-06	Event	VOEvent 1.1	WD	Williams
Sep-05	GWS	VO-Support Interface-V1.0	WD	O'Mullane, Rixon, Thakar
Nov-05	GWS	VO- Web Service Interoperability-V1.0	WD	Schaaf
Oct-05	GWS	Single Signon Authentication V1.0	WD	Rixon
Jan-06	GWS	Universal Worker Service -V1.0	WD	Rixon
May-06	GWS	VOSpace V1.0	WD	Graham, Morris, Plante
Oct-03	SD	IVOA Document Standards 1.0	REC	
Jul-05	VOQL	Astronomical Data Query Language -v1.01	WD	Nieto, Shirasaki
Jun-05	VOQL	Skynode Interface 1.0	WD	O'Mullane, Ohishi
Jul-05	VOQL	SkyNode (Queries on joint catalogs) -v1.01	WD	Nieto, Shirasaki
Mar-05	Registry	VO-Identifiers V1.10	PR	Plante
Jun-05	Registry	Resource Metadata V1.1	PR	Hanisch
Jul-05	Registry	VOResource (schema)-V1.0	WD	Plante
May-06	Registry	Registry Of Registries -V1.0	inWG	Plante
May-06	Registry	Registry Interface-V1.0	inWG	Benson
Oct-04	Semantics	Unified Content Descriptors 1+ V1.06	REC	Derriere, Preite Martinez
Jun-05	Semantics	Create vocab and tech editorial boards	Note	Preite Martinez
May-06	Semantics	Objects/Processes vocabulary	Note	Preite Martinez
Aug-04	Table	VOTable-V1.1	REC	Ochsenbein

Date	WG/IG	Standard	Status	Responsible
Aug-05	Table	VOTable-V1.2	WD	Ochsenbein
Aug-05	Theory	Theory use cases	Note	Lemson