International

Virtual

Observatory

Alliance

The IVOA in 2007: Assessment and Future Roadmap

IVOA Technical Coordination Committee

Version:

2007 May 1 this is DRAFT 2

Author:
Roy Williams

In consultation with:
IVOA Technical Coordination Committee:
Mark Allen, Francoise Genova, Bob Hanisch, Gerard Lemson, Jonathan McDowell, Reagan Moore, Francois Ochsenbein, Pedro Osuna, Ray Plante, Andrea Preite Martinez, Guy Rixon, Doug Tody, Nic Walton

Abstract:
This document is the result of a study by the IVOA Technical Coordination Committee with the intention of coordinating the IVOA Working Groups and Interest Groups. Specific objectives are:

- 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.

1. Progress since Interop 2006

In the last year, there has been considerable progress in IVOA. The Registry infrastructure is now essentially complete, and full implementation by the global VO registries is imminent. The DAL group has completed the Spectral Access protocol (SSAP) specification, as well as great progress with related protocols, all built uniformly with the Characterization data model. The collection of protocols generically called Skynode has been factored into a usable query language based on SQL92, a Table Access Protocol, and a variety of crossmatch algorithms. On the desktop, users of astronomical software are able to work easily with remote services, as well as local applications working together through standards such as Plastic. The promise of VOSpace is that storage will no longer be tied to a specific physical data location. Some national projects offer registration and certificates so that a small number of authentication credentials will have wide applicability without continual memorizing and typing of passwords. Several standards have been ratified by the IVOA Exec, including VOResource (registry record), Identifier structure for VO resources, VOEvent (for notices of transients), and a new set of UCD words.

2. IVOA Methods of Work

IVOA interoperability advances through a balanced combination of "bottom-up" and "top-down" development. Bottom-up means developing standards that address current and imminent needs characterized by science-based use cases; top-down means maintaining a vision and overall roadmap of where development is going and what capabilities to be enabled in the future. Bottom-up developers can keep the top-down model in mind so allow future enhancements without design/implementing the entire model. This balance goes on at both the executive level and the level of individual working groups.

The IVOA long-term vision is built by a sequence of short-term, incremental deliverables. Standards are sized so that they can be developed, on average, in one year (from internal WD to Recommendation). IVOA needs interoperability, but this does not imply homogeneity of the VO. In particular, projects need the ability to innovate and deliver added value that addresses the specific needs for their own community of users. We are vigilant in seeking out areas where projects are working on the same capabilities, and we evaluate whether interoperability is improved by collaborating on a common solution.

A typical scenario for the emergence of new standard would proceed as follows: A project pursues development of a new capability that meets a need of their community; the project lets other projects know about it via IVOA Interops and other meetings; the broader IVOA community recognizes the capability as generally useful (perhaps after seeing a demo); a working group takes up the task of turning the capability into a standard; finally the standardization completes.

We note that Working Groups establish focused teams to develop a particular standard. We note that Working Group chairs and vice-chairs work together to share administrative load.

3. 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.

4. 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.

5. Roadmaps for the Future

The current roadmap situation (May 07) 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. Two interoperable implementations are needed for a standard to be considered as a PR.
- **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

6. Leading Issues

(1) GetCapabilities method for Services
Many of the IVOA services are architected with a *getCapabilities* method, which should return metadata about the service. There is a debate in the DAL/VOQL/Registry communities about exactly what is returned and its format; and especially the question of whether this is different from the registry record. In other service-oriented architectures (eg OpenGIS geographical information), the registry record is very light – just a service endpoint – and the full record is obtained from the service itself. This also means that registries get records in three distinct ways: by calling the *getCapabilities* method of a service; by submitting a registry record to a registry; or by harvesting from another registry.
Recommendation: That the VO should allow this method of registration, as recommended in the VO Support Interfaces proposal from GWS. Therefore the *getCapabilities* method name will be reserved so that a service can returns its own registry record (VOResource) in the standard XML encoding. There should also be tools so that service providers can create and edit VOResource records.

(2) SOAP and REST
In the IVOA, the term "web service" generally implies either SOAP or GET/POST/REST type service protocol. The latter are simpler to understand and implement and the software is much less complex and bug-infested, and therefore preferable for simple services; however, in some cases the extra sophistication of SOAP makes it optimal. A significant advantage for SOAP services is that it is easy to create a formal interface document (WSDL), whereas this is more difficult for GET/POST/REST services (done by hand).
R**ecommendation:** A IVOA study to understand where SOAP is sufficiently advantageous and where the easier GET/POST/REST can do the job just as well.

(3) Asynchronous services
As the VO concept matures, asynchronous services are emerging, where the response to a request is not the answer, but rather a way to check on the running service, which will eventually produce the answer. There is already deployment of asynchronous services (UK-VO, US-VO, France-VO, Euro-VO), but multiple proposed standards. The GWS proposal focuses on estimating the time the job will take (*JobList/Job/Quote/TerminationTime/Quote/Phase*), and on exchange of XML documents; the DAL proposal integrates asynchrony with astronomical services through the *stageData/getData/AccessReference* attributes of the S*AP protocols. Other implementations use a protocol of *init/upload/run/monitor/destroy*. However, it is difficult to find a developer who is using a standard that is “not invented here”.
Recommendation: Assess differences in approach to async services (real or imaginary?). Assess costs and benefits of interoperability in different kinds of async services. If interoperability is desired, evaluate costs and benefits of different approaches.

(4) Rich Registries
While the IVOA registry structure is essentially complete, its mode of usage is still a topic of research. While some use small and agile registry records that refer to basic resources, others see the registry as a place to put detailed, sophisticated, dynamic summary information about a wide variety of resources. One possibility is to use other (non-registry) metadata stores. If the choice is made to use the registry, then flexibility is possible through the use of *schema extensions* – and these may be mandatory or not. For the small number of mandatory extensions, all registries must index their content, and there is a debate over how much information to include in those. There are groups that are now developing applications that use the registry to do discovery, and to do workflow planning based on more fine-grained information, and they are using registries to manage that information, assuming that their

extensions will be mandatory. The problem with this is that it puts pressure on all publishers to provide this information and on all registries to support, and therefore, curate the information. Non-mandatory schema extensions are easier: there is a working agreement that any registry must store any resource document, but need not support searching that document. However, the IVOA should still keep track of all schema extensions, even non-mandatory.
Recommendation: A mandatory schema extension must be approved as an IVOA Recommendation, with the full process ending in approval by the IVOA Exec.
Recommendation: Each non-mandatory extension schema must be explained by an IVOA Note, be approved by the Registry Working Group, and that a list be kept by that WG of all approved extensions and the corresponding description document.

(5) Registry Harvesting and concatenated XML
A problem has emerged in the last year concerning the XML documents that registries exchange in the process of harvesting each other, and this is blocking the progress to Recommendation of the VOResource standard. A set of these documents (instances of VOResource) is handled by the registry with the (false) assumption that a concatenation of valid XML documents is also valid. The problem is with the ID construct in XML, which states that such ID values must be unique. In particular, the STC schema uses these IDs to identify coordinate systems for spatial coverage, although we should say this is a general XML problem, not specific to STC. A user might write
ID="UTC-FK5-GEO" href="ivo://STClib/CoordSys#UTC-FK5-GEO" meaning the ID value can be used as an abbreviation of the referent (href value). However, if the same abbreviation is declared elsewhere in the document, the XML rules make it invalid, hence the problem with concatenating documents that all use the same coordinate system. A solution is emerging based on the following agreements (a) the ID value can and will be changed arbitrarily in an XML document without changing the essential information, and (b) this is easier to do if all ID values are easy to find in the XML; therefore (c) parsing software for the XML document must make decisions based on the referent value, not the ID value, and (d) the referent of the ID must be well-defined and cast in stone, so that parsing software can recognize it.
Recommendation: All IVOA standards and software be examined for a reliance on explicit values of the ID attribute in XML.

(6) Registry of Registries
Work has been ongoing in the Registry working group to create a "Registry of Registries", as a guide for humans to choose where to register their resources, and also so that registries can automatically choose harvesting targets. The work so far has concentrated on the difficult task of building automated, detailed compliance-checking of candidate registries.
Recommendation: A precursor to the RofR system should be made at ivoa.net, in the form of a simple HTML list of registry URL and endpoints, with contact information.

(7) Data Models and *utype*s
The concept of "*utype*" was defined in the IVOA as a response to the fuzzy nature of the UCD descriptor: if a quantity has a *utype*, then it must be part of a specific data model. Proper *utypes* would allow queries to be built independent of the underlying database structure ("where STC.coords.FK5.RA between 300 and 302"), and would provide a strong framework for parameter-based queries ("http://.....? STC.coords.FK5.RA = 300 &..."). However, many of the data models in use in the IVOA have XML representation only, and do not have representation as a hierarchy of *utype* values. We note that the syntax of *utypes* is not well defined in the IVOA, and also that in simple cases the *utype* can be cleanly derived from the Xpath representation of an XML element, so this should be a straightforward matter.
Recommendation: The syntax of *utype* and its namespaces should be well-defined. Just as with UCDs, there should be services to find relevant data models and their *utypes* from search words, and there should be services to trace a given *utype* back to its precise meaning.

(8) IVOA Recommendation: what is an "implementation"?
The IVOA requires two "interoperable implementations" as part of the requirements for a proposed standard to become a Recommendation, but does not define exactly what is meant by this. Some see this in a weak sense: maybe some piece of prototype software that exercises some part of the proposed standard, or even simple storage of a relevant document without even parsing it. Some see a much stronger requirment: that real users prove the standard in real science code.
Recommendation: The IVOA Documents and Standards WG should collect the will of the IVOA Exec and encapsulate the meaning of the words "interoperable implementations".

(9) Space-Time Coordinates
This large and comprehensive working draft has become a de facto standard in the IVOA through multiple implementations, and yet it is not yet a Recommendation. The IVOA should take firm action on this matter to resolve the status of STC. While there are several software packages that use STC, none of them exercises *every* part of the proposed standard. Further, there is often complaint from implementers about the complexity of STC -- countered by the contention that astronomical coordinate systems are complex by nature. What astronomers want in this area is **both** assurance that full rigor and precise coordinates are available in the IVOA; **and** the release from complexity when that full rigor is not deemed necessary by the astronomer.
Recommendation: In addition to STC, there must be a simpler system for everyday use, with mappings to full STC well-defined. It is a matter of defaults. For example if the information in the simple system is just RA and Dec numbers, this can map to the FK5 system with reference point at the barycenter of the solar system and the epoch 2000.0. Regions that are disks and RA/Dec intervals should be expressible in just a few characters.

(10) Regions of the sky
The VO projects use multiple description schemes for a subset of the celestial sphere: a disk on the sky, an aligned box (RA and Dec limits), polygons and ellipses as well as boolean operations on half-spaces and expressions as unions of different sized pixels. The IVOA uses region specifications in multiple ways: for describing coverage of a dataset, as part of a parametrized query, as part of an ADQL query, etc. Decisions about this matter seem to be taking place in several overlapping projects and working groups, and an IVOA-wide approach would help. In particular, we note that components of STC are used by some projects, but others opt for a simpler representation scheme, eg CIRCLE(200,30).
Recommendation: The various descriptions of regions should be compared with the various uses and reasons that these descriptions are used. How are region descriptions parsed and evaluated in the context of real scientific use cases? What level of complexity is actually needed to do science?

(11) Application Messaging
The Applications WG has been responding to the increasing popularity of the Plastic protocol, that allows desktop applications to exchange messages. The initial discussion has explored the wide range of intended uses for messaging between applications in the VO. Two different types of messaging have been identified (i) PLASTIC style messages which deal with high level data objects like images and tables which can be sent with a loosely defined message as in *'here is an image*'. (ii) messages which require a more precise description, such as commanding other tools to execute a parametrized task. It is universally agreed that messaging protocols should be decoupled from the semantics of the messages. Much of the debate has been on how and whether both types of messages should work within a common framework. A promising general framework has been proposed which consists of a separate message container, and message content model elements. An immediate concern has been the cost versus benefits for PLASTIC

style messaging in such a general framework, and there is a strong desire to ensure PLASTIC style interoperability in real implementations through any transitions.

(12) VO interoperability with popular software
Most astronomers do most of their work with software packages like IDL, IRAF, DS9, MIDAS, Sextractor, etc. It is highly desirable that these be interoperable with the VO framework. Applications messaging is one aspect of this, but a more complete consideration of what this means is required.
Recommendation: The Apps WG shold prepare a report of VO interoperability with astronomy software packages and environments.

(13) Bundling of VO software
Bundling of astronomy software such as the Scisoft and ex-Starlink collections provides a convenient way of distributing many packages at once to ease the burden of installation. Bundled distributions of VO software would assist in up-take of VO tools, and we note that Scisoft VII will contain a selection of VO software.
Recommendation: The list of VO Applications maintained on the (publicly editable) Apps WG wiki pages serve as a place for Applications to be visible for parties compiling collections of VO tools.

(14) Interoperable Security: Security and authentication is being implemented in several new efforts. The Astrogrid (UK-VO) project has built a sophisticated workflow system for asynchronous computations and is adding authentication; a complementary project from the US NVO project is exploring the idea of "graduated security" for giving community access to high-performance computing. While the IVOA has a mature Single-signon standard for security, using X.509 certificates, there has been little discussion of which VO projects are issuing certificates and the levels of authentication taking place, and which VO projects will accept certificates from which other projects.
Recommendation: Creation of an IVOA listing of certificate authorities in the national projects, how to get a certificate from each, and what can be done with the certificate.

(15) Units: Most scientific quantities carry units, and data returned from IVOA services should also carry explicit unit information when not clear implicitly. Units should follow the IAU recommendation[1], which follows the SI convention. When a user makes a query based on a quantity, units can either be user-defined or fixed. In the former case, the user has the freedom to express the quantity in arbitrary units (eg. *calories per square furlong per hour*!), or an enumerated choice (eg. *Angstroms OR nanometers*). In the case of fixed units, the data model of the query is bound to specific units (eg *all angles must be in decimal degrees*).
R**ecommendation:** A study by the Data Model Working Group of how units are used in IVOA views and services, where it would be appropriate to simply fix the units, and where it is necessary to allow freedom of choice, distinguishing between unit choice in the user interface and in the back-end services. In the latter case, the report should also recommend on how unit conversion is implemented: who is responsible and the nature of the software.

(16) IVOA Newsletter
Recommendation: The global VO community would be well-served by an IVOA newsletter, including announcements from national projects and working groups, events, press coverage of VO issues, etc.

Table 1: **IVOA Working Groups and Interest Groups**

[1] Recommendations Concerning Units, http://www.iau.org/Units.234.0.html

Working/Int. Group		Chair and vice-Chair	Current priorities
Applications	WG	Mark Allen Mark Taylor	Various application news. Application messaging standards.
Data Access Layer (DAL)	WG	Doug Tody Markus Dolensky	Spectral Energy Distribution (with DM). Simple Spectral Access, Level 2 Image Access (datacube), Characterization and “generic dataset”.
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 MIreiille Louys Anita Richards	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 Rob Seaman	Production implementations and community partnerships. Prototyping new features. Event transport. Event Semantics WD 1.1 and schema.
Grid-Web Services (GWS)	WG	Guy Rixon Matthew Graham	Security, trust, single sign-on. VOSpace. Asynchronous services. Support interfaces for services: metadata extraction, availability reporting, user groups, service logging.
Query Language (VOQL)	WG	Pedro Osuna Yuji Shirasaki	Astronomical Data Query Language (ADQL) as XML and script. SkyNode Interface methods. Integration with DAL
Registry	WG	Ray Plante Aurelian Stebe	Resource Metadata, semantics and schema. Service Interfaces. Registry of registries Registering general services and applications. Query languages for the registry.
Semantics/UCD	WG	Andrea Preite-Martinez Sebastien Derriere	Updating and agreeing UCD list. Workflow for changes to list. Role of ontology. Standard vocab for Process/Objects
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 Herve Wozniak	Data Modelling and Formats (Lemson et al); Access Protocol – N-body and mesh simulations Semantics and UCDs for Theory (Shaw et al).

Table2: **IVOA WG Roadmap May 2007**

Date	WG/IG	Standard	Status	Responsible
May-07	Apps	Application Messaging (SAMP)	Note	Allen, Fitzpatrick, Taylor, Taylor
Feb-07	DAL	Simple Cone Search V1.0	PR	Plante
May-07	DAL	Simple Numerical Access	inWG	Lemson
May-07	DAL	Spectral Line Access V1.0	PR	Salgado, Osuna
May-07	DAL	Simple Spectral Access V1.0	PR	Tody, Dolensky
May-07	DAL	SIA-Level2 (cubes etc)	inWG	Tody, Bonnarel
May-07	DAL	Table Access Protocol	inWG	Osuna, Tody
Jun-07	DAL	Simple Image Access V1.0	PR	Tody, Plante
Aug-05	DM	Atomic Line Lists-v1.0	WD	Dubernet, Osuna
Jul-06	DM	Spectrum-v1.0	WD	
Jul-06	DM	Space Time Coordinates-V1.3	PR	Rots
Oct-06	DM	Characterisation-V1.0	WD	Bonnarel, Louys
Dec-06	DM	Spectrum-v1.0	WD	
May-07	DM	Spectral Energy Density-V1.0	WD	McDowell, Tody
Mar-07	DM	Source Catalog Model	PR	
2008	DM	Dataset Registry Resource		
Jul-05	Event	VOEvent 1.0	WD	Seaman, Williams
Oct-06	Event	VOEvent 1.1	REC	Williams, Seaman
Jun-07	GWS	VO Basic Profile V1.0	PR	Schaaf
Jul-07	GWS	VO Support Interfaces	PR	Rixon
Jul-07	GWS	VOSpace V1.0	PR	Graham, Morris, Plante, Rixon
Jul-07	GWS	Single Signon Delegation Services V1.0	WD	Rixon
Aug-07	GWS	Single Signon Authentication V1.0	PR	Rixon
Aug-07	GWS	Single Signon Community Services V1.0	WD	Rixon
Sep-07	GWS	VOSpace V1.1	WD	Graham, Morris, Plante, Rixon
Dec-07	GWS	VOSpace V2.0	inWG	Graham, Morris, Plante, Rixon
2008	GWS	Harvesting logging data	WD	Thakar
Oct-03	SD	IVOA Document Standards 1.0	REC	Hanisch
Jul-06	VOQL	Astronomical Data Query Language -v1.04	WD	Nieto, Shirasaki
Jul-06	VOQL	Skynode Interface 1.02	WD	Nieto, Shirasaki
Dec-06	VOQL	SkyNode Extensions	WD	Nieto, Shirasaki

Date	WG/IG	Standard	Status	Responsible
May-07	VOQL	Astronomical Data Query Language 1.1	WD	Osuna, Tody
May-07	VOQL	Table Access Protocol	WD	Osuna, Tody
Jun-06	Registry	Resource Metadata V1.1	REC	Hanisch, Linde
Jun-06	Registry	VOApplication V1.0	WD	
Jun-06	Registry	CEAApplication V1.0	WD	Harrison, Linde
Jul-06	Registry	Registry Of Registries	inWG	Plante
Nov-06	Registry	VO-Identifiers V1.1	REC	Plante
Nov-06	Registry	VOResource V1.0	REC	Plante, Linde
Jun-07	Registry	Registry Interface V1.0	WD	Benson, Linde
Jun-07	Registry	VODataService v1.0	WD	Plante, Linde
Aug-05	Semantics	Unified Content Descrptors, V1.10	REC	Derriere, Preite Martinez, Williams
Dec-05	Semantics	The UCD1+ controlled vocabulary Version 1.11	REC	Derriere, Preite Martinez
Jun-06	Semantics	Maintenance of the list of UCD words	REC	Derriere, Preite Martinez
Apr-07	Semantics	The UCD1+ controlled vocabulary V1.23	REC	Derriere, Preite Martinez
Oct-07	Semantics	Ontology of astronomical object types: A use-case	Note	Derriere, Preite Martinez, Richard
Aug-04	Table	VOTable V1.1	REC	Ochsenbein
Aug-06	Table	VOTable V1.2	REC	Ochsenbein
Oct-06	Theory	UCD Extensions for Theory	Note	
Oct-06	Theory	Simple Numerical Access Protocol V0.1	WD	Lemson