



International

Virtual

Observatory

Alliance

Proposal for a *Simulation Database Standard*

Version 1

Note 2008 July 11

This version:

1-20080711

Latest version:

<http://www.ivoa.net/Documents/latest/NOWHERE>

Previous version(s):

Author(s):

Gerard Lemson
Hervé Wozniak
Laurent Bourgès
Rick Wagner
Claudio Gheller

Abstract

We propose to the IVOA Exec and TCG that the “Simulation Database (SimDB)” specification that has been worked on in the IVOA Theory Interest Group is moved to the recommendation track. Here we describe this specification *to be* in not much detail, focusing on those features that require special attention in the evaluation of this proposal.

We believe that further evolution of this specification requires a multi-disciplinary (i.e. multi-working group) approach. To resolve the problems this poses for the standard, single-WG approach in the IVOA, we present a number of ways in which the recommendation track may be implemented.

Status of This Document

This is a Note. The first release of this document was 2008 July 11.

This is an IVOA Note expressing suggestions from and opinions of the authors. It is intended to share best practices, possible approaches, or other perspectives on interoperability with the Virtual Observatory. It should not be referenced or otherwise interpreted as a standard specification.

A list of [current IVOA Recommendations and other technical documents](http://www.ivoa.net/Documents/) can be found at <http://www.ivoa.net/Documents/>.

Acknowledgements

We thank Christophe Arviset, Ugo Becciani, Severin Gaudet, Mireille Louys and Patrizia Manzato for discussions on this topic.

Contents

1	Introduction	2
2	Summary	3
3	SimDB and working groups	5
4	Recommendation track	7
	References	8
	Appendix A: some details on SimDB	8

1 Introduction

We here describe a proposal for an IVOA standard, the *Simulation Database* (SimDB). Work on this standard started under the header *Simple Numerical Access Protocol* (SNAP) in the theory interest group since the Victoria interoperability meeting, 2006. Recent developments have made us decide to

split SNAP in two separate tracks, SimDB and SimDAP, which stands for *Simulation Data Access Protocol*. The main part of this *Note* deals with SimDB only: what it is supposed to be, what its current state is, and what further work is needed and the possible organisation of that work.

Work on this specification has been organised via a googlecode SVN repository in the *volute* project originally created by Norman Gray for the Semantics Working group. The SimDB project in particular can be followed under <https://volute.googlecode.com/svn/trunk/projects/theory/snapdm>. In particular follow the SimDB_Note.html document under <https://volute.googlecode.com/svn/trunk/projects/theory/snapdm/doc/note> which contains the current status of the specification document [2].

SimDAP deals with extracting subsets of data from simulation results that have been discovered through a SimDB. Its dependence on SimDB is through its dependence on aspects of the SimDB *data model*. The progress of the SimDAP specification can be traced also via the googlecode volute project: <https://volute.googlecode.com/svn/trunk/projects/theory/snap> (see also [3]).

We think the SimDB effort has evolved far enough that it can be moved onto the recommendation track. After summarising its main features, we identify some issues related to this next step, namely that it is hard to find a single working group that could be responsible for tracking all aspects of the specification. We provide a number of possible solutions to this, but leave it to the TCG and the Exec to decide on this. In the mean time the current group of developers will keep working on this, with the goal of having working prototype/reference implementations available by the October 2008 interop.

2 Summary

SimDB is a specification for an online web service providing access to a repository storing metadata about numerical computer simulations of astrophysical systems and related resources. It might be also be called *Simulation Registry*, or *Simulation Portal*. Currently the simulations are still (following the original SNAP approach) supposed to be those that produce a representation of 3+1D space. In later versions this restriction is likely to be relaxed.

A SimDB is supposed to be used to *discover* simulations together with web services providing access to them. The more detailed specification of these services is the goal of the SimDAP-specification.

The following are the normative aspects of SimDB:

1. SimDB is based on a (logical) data model, fully specified in UML2.
2. From the UML data model we derive physical models for use in their respective SimDb service contexts:

- a. A relational database schema expressed according to the TAP specification.
 - b. An XML schema, defining valid XML documents containing SimDB meta data descriptions for use in messaging.
 - c. A set of UTYPEs identifying elements of the model in case this model is to be expressed in VOTables or other non-SimDB-standard representations. One context for these is as the results of the ADQL queries.
 - d. A human readable HTML document describing all the individual model elements in detail.
3. These physical representations are to be used in the service interface specification of SimDB instances. These are
- a. An ADQL-based querying of the metadata repository as a relational database. Our hope is that would be based on the upcoming Table Access Protocol (TAP). Barring that, we would define a minimal service interface consisting of a parameterized GET or POST to pass the ADQL, and a response formatted in a VOTABLE (similar to the current SkyNode).
 - b. A RESTful web service interface, using standard HTTP methods (GET, PUT, POST, DELETE, etc.) to provide mechanisms for maintaining the actual entries in a SimDB. SimDB resources would be uploaded, retrieved, or modified using the XML format defined by the XML Schema derived from the data model.
 - c. Possibly an OAI-PMH compliant publishing interface, to allow harvesting of SimDB records. Like the Registry, this would permit SimDB instances to acquire records published in other SimDBs.

The precise specification of the service aspects is still relatively open, but will follow standard (IVOA) practice where applicable. We expect that we do not need full fledged asynchronous, paging services etc, as we are dealing with metadata repositories which are expected to be relatively small. Nevertheless this aspect will be formally defined.

This concludes the normative/standard aspects of the SimDB as far as it has currently been designed. We do want to mention the following aspects of the SimDB effort that will simplify much of the further work:

1. We have XSLT scripts that derive the physical models directly from the UML model according to predefined mapping rules.
2. We also derive Java classes with JPA and JAXB annotations to make it easy to implement a SimDB from the specification.
3. From this we are developing a full SimDB implementation using code generation from the UML model only.
4. We propose an implementation path to transform an existing, "legacy" relational database containing simulation metadata to the SimDB specification.

The following groups have committed to creating a reference implementation of a SimDB, some of these are meant to be ready by the October 2008 interoperability meeting.

- GAVO (Gerard Lemson)
- UCSD (Rick Wagner)
- VObs.It (Ugo Becciani, Patrizia Manzato)
- VO-France (Laurent Bourges, Franck LePetit)

3 SimDB and working groups

This specification overlaps the domains of multiple IVOA working groups. This causes some problems for the promotion of the effort to a recommendation track. The IVOA approach [1] mandates that only working groups can promote specification on this track. In the current case there is no one working group that has a mandate covering all aspects of the proposed specification. This will be discussed more in the next section.

But in whatever way the responsibility for further evolution of SimDB is to be organised, the proposal as is should be investigated by various WGs. Furthermore there are open issues that still have to be resolved that require input from a number of working groups. Here we give for each of the relevant working groups some points that we feel are their domain.

Theory [TIG]

- We need to confront scientists with our ideas and extract feedback. This includes defining use cases and the content of the data model.
- Promote take-up by these simulators.
- Maintenance of SimDB and further evolution to incorporate other types of simulations remain domain of TIG (or doesn't it?).

Data Modeling Working Group [DM]

- We propose a specific UML profile for use in SimDB data model.
- We must discuss the relation to other data models further, e.g. Characterisation.
- We propose a way to incorporate semantic vocabularies in the data model.
- We propose a mapping of our UML data model to relational and XML schemas
- We propose an algorithm for deriving UTYPEs from the UML data model.

Data Access Layer [DAL]

- Originally what is now called SimDB corresponded to the queryData part of the DAL-like SNAP. Should it still be that, or is it more registry like?

- If SimDB remains S*AP like, can we nevertheless support a centralised repository to implement the *queryData* part for many distributed simulation archives and access services?
- We are proposing to use ADQL in the *queryData* part of the protocol.
- We are proposing to use TAP to describe a fixed, global relational data model.
- Modelling of applicable data access web services for accessing simulation results.

Resource Registry [Reg]

- SimDB acts as a registry, but isn't one (or is it?). Would be good to have similar solutions for service interfaces etc.
- SimDB data model has similarities with Registry data model, e.g. introduces SimDB::Resource and SimDB::Curation, which are similar to, but not identical with similar components in Registry model.
- The data model is rather normalised, indicating that we need references between resources and sub-elements of these. To this end we propose an upgrade of the concept of IVO identifiers.
- Should a SimDB be stand alone, or should it be possible to have relations between SimDB/Resource-s in different SimDB-s? The answer to this has repercussions for the XML and RDB representations. It may have repercussions for the functioning of a SimDB in case it is supposed to be stand alone (mirroring through harvesting of external resources may be required).

Semantics [Sem]

- We need to use common vocabularies to identify various elements in data model. Semantics is the WG that defines these.
- We propose a way to link the data model to such vocabularies in the UML profile. Comments from both Dm and Sem are desired.
- We need automated ways to retrieve such vocabularies and their elements.

VOQL

- We want to use ADQL to query the data model and will require ADQL parsers/translators. Construction of this we believe is an effort of the ADQL tiger team?
- We are using ADQL in a way that is possibly different than foreseen, querying highly normalised meta-data databases. We want to make sure ADQL is appropriate for this.
- See the entry on UTYPEs under DM. VOQL is interested in these as well.

4 Recommendation track

As we have hopefully made clear, we need input from several WGs on the current state and further development of SimDB. In particular, if we want to promote SimDB to the recommendation track, we need to be able to start writing working drafts of the SimDB note [2] and move this through the recommendation process. Under the current rules of the IVOA [1] an interest group can not do so, and it might in any case not be appropriate to bypass formal interaction with the relevant WGs.

The following are four proposals that have been made in the past to deal with this issue:

1. **Choose 1 WG.**

It has been suggested that DM could take the responsibility, as SimDB has a data model at its core. However it could just as well be Registry, as that group has more experience with defining and maintaining services ala SimDB, including defining complex data models.

2. **Turn TIG into TWG.**

It has been suggested often in the past to turn the theory *interest* group in a theory *working* group.

3. **Divide SimDB in two pieces: SIMDB/DM + SimDB/Service.**

SimDB/DM would be a specification for the data model only, SimDB/Service for the service interface to a SimDB implementation.

4. **Create a problem specific *focus group* with promotion privileges.**

A focus group (FG, we try to avoid the name “tiger team”) is formed from participants of all relevant WGs and IGs, but not necessarily under the header of a single WG. It is formed and mandated by the IVOA Exec for a particular purpose, i.e. standard specification project. Chair of the FG has same responsibilities as chair of a WG.

The first two of these proposals have as benefit that they fit within the current formal structure of the IVOA standards process. However we feel they do not do justice to the truly multi-WG nature of this projects and other possible projects like this (e.g. UTYPE, Units), in that they do not mandate or organise the relation to these other WGs.

In a sense, the approach to make the TIG a working group is similar in nature to having a separate working group for the resource registry, which itself has defined a data model and defines access services to registry implementations, in short has had overlap in its efforts with other WGs. The TWG could therefore simply do what they want and wait till the RFC to get formal interaction with the other WGs. We do however not think that that is the preferred method, instead we want there to be a solution that makes the interaction with the relevant working groups more immediate, which is what the fourth proposal might accomplish.

The idea behind the proposal #3 is that at least SimDB/DM could find safe haven in the DM WG. There are various problems with it though. First, even at the level

of the data model alone SimDB has to cross into the domain of the Semantics working group, as we need to be able to describe objects explicitly using semantic vocabularies (or some other appropriate form of ontology-like results). Second, this “solution” merely pushes off the problem as a solution must be found for the SimDB/Service specification, which is still a mixture of WG domains.

But the more severe objection is that neither of the two specifications would really stand on its own. The SimDB data model is influenced by the use cases for which it was designed, namely that it provide a description of simulation results of sufficient richness to discover them and be directed to web services for accessing these results. It is not the only possible model for describing simulations and certainly not the most comprehensive model. Similarly the SimDB service interface is very much influenced by the actual character of the SimDB/DM. This is relatively complex and requires more sophisticated techniques for implementing services around that the typical S*APs. It leads one more closely to a registry-like service. For a simpler model this would not be required.

Finally, there is an alternative version of the fourth proposal, which says that

4b. The TCG defines a problem specific focus group from its members.

This focus group would oversee the work on a particular multi-WG project and ensure that evaluation steps are synchronised between the working groups.

In this approach, the specification itself would likely still be owned by a working group, but when creating new versions of a working draft not only that WG would have to agree, but also all the other ones. It would be the responsibility of this TCG-focus group to ensure that this happens in a timely fashion.

Summarising, our preference is an approach using a focus group (proposals #4 or #4b) for moving SimDB towards RFC and beyond is preferred.

We leave it up to the Exec and TCG to further decide on this issue.

The SimDB proposal being completed and the reference implementations being in operation by the next fall INTEROP, we expect having a recommendation process defined by October.

References

[1] R. Hanisch, *IVOA Document Standards*,

<http://www.ivoa.net/Documents/latest/DocStd.html>

[2] G. Lemson *et al*, *Simulation Database (SimDB)*,

<http://code.google.com/p/volute/source/browse/trunk/projects/theory/snapdm/doc/note/SimDB-note.html>

[3] C. Gheller *et al*, *Simulation Data Access Protocol (SimDAP)*,

<http://code.google.com/p/volute/source/browse/trunk/projects/theory/snap/SimDAP.html>