

Small Theory Data in the Virtual Observatory

Jonathan C. McDowell

Smithsonian Astrophysical Observatory, 60 Garden St, Cambridge, MA
02138

Abstract. The integration of large theoretical simulation archives with the VO has been widely discussed. I suggest that it is also important to include smaller theoretical datasets and functional relationships in a structured way, and outline some possible conventions.

1. Introduction

In this article, I address the issue of resource discovery for tabular and functional theoretical and phenomenological results such as extinction laws, luminosity functions, isochrones, and distance indicators. A structured extension of the CDS concept of UCDS could make tabular data of this kind easily available not only to astronomers but also to interoperable software.

I also discuss metadata for simulations by drawing an analogy with X-ray spectral analysis, a domain in which complex new theoretical models have been rapidly integrated with the standard data analysis tools via a simple parameterized-function description. This paradigm can easily be extended to image simulations.

2. The Virtual Astrophysics Library

When I was a theorist, I spent a lot of time retyping tables from the ApJ and coding small equations from papers as subroutines; and best of all, using a ruler and pencil to digitize xeroxes of graphs. The Virtual Observatory (VO) can really help here.

My challenge to the reader is: every time you see a graph or a histogram in the ApJ or on astro-ph, I want you to ask yourself:

- How would I encode this in the VO?
- How would I locate this info in the VO?

I propose the Virtual Astrophysics Library, an on-line collection of astronomical relationships – a place to find different versions of the extinction law, the initial mass function, the Tully-Fisher relation, etc., either in tabular or subroutine form.

The near term goal is to provide coherent access to useful snippets of data, saving retyping and recoding. A longer term goal is on-the-fly manipulation of VO observational data. As presently envisaged, the VO will allow the user to say “give me this image”; I believe we should be able to say “give me this image,

dereddened with an LMC extinction curve and K-corrected to redshift 2 using Smith’s spectral energy distributions and the following cosmological model..”

2.1. Step One: Small Theory or Phenomenology Tables

Much of this functionality can be encoded in fairly small lookup tables, which are easy to convert to self-describing (e.g., VOTable¹) form. The trick is accessing (i.e., indexing) them. For example, an extinction law is a simple function (Figure 1a), but it is not analytic: rather, it is a mixture of measured and calculated values. One can consider it as a function of three parameters: wavelength, possibly metallicity, and version (Seaton et al. 1979; Morrison & McCammon 1983, etc.) Note that not all versions cover all possible wavelength ranges, so we must provide a standard which allows the function publisher to describe not only the function arguments but their allowed values.

We may wish to provide a ‘VO canonical’ version combining our favorite ones to cover all wavelengths (e.g., as done by Joachim Koppen’s extinction web service at astro.u-strasbg.fr), although this raises problems of editorial authority. It is probably better to let such attempted syntheses be published on an equal basis with the individual fragments, and provide a mechanism for VO users to specify their defaults in a configuration file.

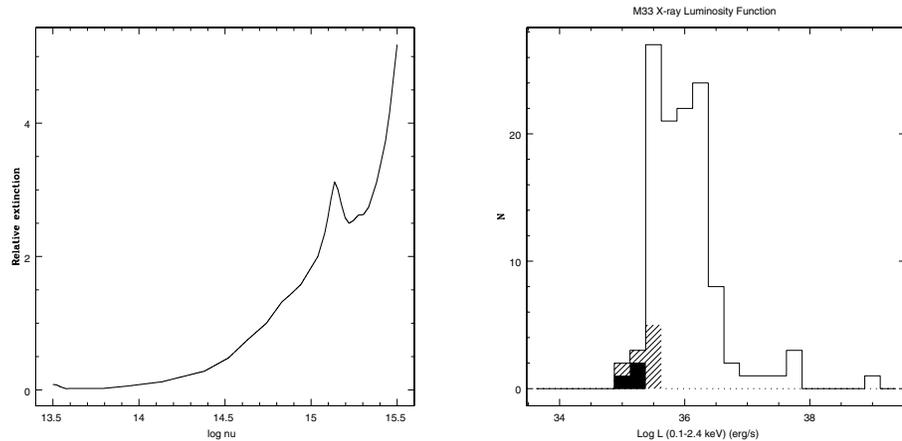


Figure 1. (a) A sample extinction law. (b) Luminosity function of X-ray sources in M33.

2.2. A Coordinate System for the Space of Concepts

The simplest functionality we should provide is an easy way to find all the extinction laws published to the VO. I would argue that a Google-like search is unsatisfactory, and we should have a structured way to find such information. When we are using the VO to search for information on specific objects, we can use a single, standard coordinate system (J2000/ICRS) to locate them. This

¹<http://cdsweb.u-strasbg.fr/doc/VOTable>

coordinate system has problems as an index (for example, it breaks down for solar system objects and components of binary stars) but it solves the bulk of the problem. There is no comparable coordinate system for theoretical concepts and models, although the UCD (Unified Content Descriptor) labels in use by the Strasbourg CDS (Derriere et al. 2002) provide a first step toward such an index. By having unique identifiers for concepts they are able to meaningfully cross-match the contents of a large number of catalogs.

In the case of celestial coordinate systems, even though J2000 is the default, the VO will support different views of the sky (equatorial, galactic, ecliptic...); it is even more important to do this in theory space – we must provide a framework for indexing concepts and at least one default example, but we mustn't impose a single view of physics which might restrict the questions the user could ask. An obvious way to store the extinction law would be under 'physics; radiation; opacity; interstellar'; but an equally valid choice might be 'physics; galaxies; diffuse material; spectrum'. The Strasbourg group have emphasized the usefulness of having a unique identifier for a single concept, but perhaps it's enough to ensure that such alternate formulations are easily and automatically mapped to each other. Defining these 'concept coordinate systems' is a key technology needed for the VO, and we will have to come up with it in the near future.

Providing a way to find and download tables of this kind allows us to take more sophisticated steps. The next step is to provide the table as a web service; a user would send a spectrum and an extinction value and get back a dereddened spectrum. Rather than provide a separate web service for each such physical problem (extinction, cosmology, etc.), it should be possible to code a single web service which could apply any tabulated data, while providing a check that the input data are the right physical quantities by checking they match the correct unique tags (which I will refer to loosely as UCDs).

The final step is to register the service with the VO query language. The existence of the UCD-like tags will allow the query language to chain the tables together, allowing the user to ask questions like: 'Give me all galaxies in the Smith catalog whose dereddened B magnitudes are brighter than 14, using reddening values from the Jones sky map.' The VO now knows about reddening at a basic level.

2.3. Rawer Datasets (Histograms and Scatter Plots)

Histograms and scatter plots are also really just tables. Consider the example of the SN acceleration Hubble diagram: a scatter plot of magnitude versus redshift. This is equivalent to a catalog of objects with two columns. Although the individual objects are in fact astronomical objects with RA and Dec values, this is no longer relevant – we are idealizing them as samples from a theoretical magnitude-redshift space. Therefore, the diagram (or rather the table it represents) will be indexed in physics space, not celestial coordinate space, perhaps as 'physics; cosmology; expansion; Hubble Diagram; data'. Predictive curves would be stored 'close by'.

Another example: the luminosity function histogram of X-ray sources in a galaxy (Figure 1b). The corresponding table would have metadata linking it back to the original catalog used to make the histogram, and forward to the broken power law fit used as an idealization of the data.

2.4. Publishing and Archiving Subroutines

Much of the simple machinery used to integrate tabular theory heuristics into the VO can be extended to handle code. A subroutine which requires no interaction and has a consistent set of input and output arguments (numbers, strings or files) is equivalent to a table with an infinite number of rows in which each argument corresponds to a table column. In particular, we can use the same kind of metadata used to define table columns to define the subroutine input and output arguments. The UCIDs can be used to ensure that the arguments are of the appropriate astrophysical type and not just the appropriate computer data type. One could even define an enhanced VOTable XML format with a stream parameter of type ‘external code’.

3. Big Theory Data: Metadata for Simulation Archives

Most discussion of theory data for the VO to date has focussed on the large theoretical simulation datasets, but the community has not yet established common metadata conventions for describing these datasets, or even decided what information is important enough to be recorded. We must **soon** decide how to encode (XML, FITS header) these metadata in a standard way.

I suggest that the X-ray spectral fitting community has relevant experience which can guide our thinking. In X-ray astronomy, our inability to deconvolve the instrumental spectral response has driven us to parameterized fitting of theoretical model spectra folded through the instrument simulator. The standard packages (XSPEC, Arnaud 1996; Freeman et al. 2001) share a large code base of 1-dimensional spectral simulation codes contributed by the user community. The key idea is that the simulation is described only by its parameters; the user must understand the scientific algorithm used by the simulation by reading text documentation – all the computer needs to know is a unique identifier for the simulation and the parameters that need to be fed to it. A second key idea is that all parameters are equal (in the sense that they are described in the same way), even though some are physical (temperature, abundance, density) and some are control-related (method name, number of iterations). A similar approach may be useful to describe particle simulations.

Acknowledgments. This project is supported by the Chandra X-ray Center under NASA contract NAS8-39073.

References

- Arnaud, K. A. 1996, in ASP Conf. Ser., Vol. 101, *Astronomical Data Analysis Software and Systems V*, ed. G. H. Jacoby & J. Barnes (San Francisco: ASP), 17
- Derriere, S., Boch, T., Ochsenbein F. & Ortiz P. 2002, in “Toward an International Virtual Observatory” (Springer-Verlag).
- Freeman, P., Doe S., & Siemiginowska A. 2001, SPIE 4477, 76