Astronomical Data Analysis Software and Systems XIII
ASP Conference Series, Vol. 314, 2004
F. Ochsenbein, M. Allen, and D. Egret, eds.

Electronic Journals as Databases

András Holl

Konkoly Observatory, Budapest, Hungary

Abstract. The Information Bulletin on Variable Stars is a bulletin fully available in electronic form. We are working on converting the text, tables and figures of the papers published into a database, and, at the same time, making them accessible and addressable. IBVS Data Service will provide information on variable stars — like finding charts, light curves — and will be VO compatible. Other services could link to individual figures, data files, etc. this way.

The Information Bulletin on Variable Stars is a small journal publishing papers in the field of stellar variability research. It is now available electronically, from the first issue of 1961, in ASCII text or LaTeX source format as well as PostScript (Holl 1988).

1. Searching for information

Users of IBVS can either find papers by looking at the tables of contents, by using the simple string search facilities offered by the IBVS website or using ADS and SIMBAD.

Although locating papers using bibliographic metainfo (author, publication year, journal, etc.) works well, finding specific pieces of information – like an identification chart for a variable, recent ephemeris for another, or all the published minima (info in the paper body) is a problem. Experience / experiments show that simple search facilities are not good at this. Using key combinations like:

"TZ Ari" + "finding chart"

will either return a number of other articles or it may miss the paper containing the chart altogether. Popular search engine image finding results are similar.

To get better results we must be able to find specific information in the body of a paper. Though software with more intelligence built in might help, we think the solution is to store the papers in a structural/semantic format (O'Donnell, 1993), achieved by metainfo and mark-up. While marking up archival material is a tedious job, it can be done with small journals, like IBVS. We have started working on building a figure metainfo database. For the future, a carefully constructed LaTeX style file is the solution (IBVSephemeris, IBVSfigKey, IBVSdataKey macros). Using this metainfo better search forms can be built.

2. Structural and semantic parts of a paper

It is often desirable that parts of a paper should be accessible and addressable. As the case of Agosti (2003) shows, figures should be re-usable, one should be able to compile a catalog using figures from published papers, linking to the source.

IBVS intends to make specific parts of its papers addressable and accessible. So, for example, GCVS can refer to a finding chart, or ephemeris, using a link to the given piece of information. The chart will reside on the journal's server, the figure can appear on a GCVS result page, together with proper reference to the source. For addressing, standards should be established based on existing ones like the ASU (Albrecht et al., 1996) and the bibcode (Schmitz et al., 1995).

3. IBVS as a database

IBVS has accumulated large amounts of information. From a different viewpoint, it can be considered as a database. Carrying further the example used above, IBVS has a large collection of finding charts – and light curves, ephemerides, observed minima times, observational data files, etc. These fitted with metainfo and marked up, we can compile an atlas of finding charts, or lightcurves. And, by using IBVS facilities, one might collect all minima published in the journal on a given eclipsing binary (Holl, 2002).

The quantity of such information in IBVS is not always large enough to make a useful catalog, and the material is certainly not homogeneous. We will provide such a catalog for the users of IBVS, but it is really intended to be used by other services. Catalogs might pick and use charts or other pieces of information they want, and omit those not relevant (or good enough) for them. On the other hand, other services might put together information from different sources – like IBVS and AAVSO. So services like Aladin, OASIS (Good, 2003) or SIMBAD, might give flags to variable objects with further external info, like light curves. The problem of name resolution has to be be solved, for which we intend to use Sesame provided by CDS.

The more metainfo is needed – while papers have to be prepared by the authors – the more difficult the task of the authors will be. A possible solution for this problem would be to provide tools for the authors to prepare and/or check their manuscripts. An experimental tool is described in Holl, 2003.

4. Journals and observational data

Data on which IBVS papers are based often come from small observatories. Such data are usually not available like data from big surveys. This data is mostly photometry – small tables in the form of text files (but might be spectra, etc. occasionally.) It is desirable to make those data available along with the papers. IBVS does allow electronic-only publication of such data files. Though the number of such files is small yet, it is expected to grow. It might reach the size of a database used on its own merit. A more probable scenario, though, is that it will be used through the Virtual Observatory. There are many variability surveys, offering photometry. The future user would like to gather data from

all available sources, using VO techniques. So IBVS will present observational data in VOTable format.

Data files will be supplied with metainfo, including keywords. We will try to standardize formats to a certain level – but still we will require authors to submit plain text files. These will be converted to VOTable format automatically. Manuscript submission tools will have VOTable output too.

5. Summary

We will introduce the IBVS Data Service, which will be intended primarily for other service providers, making IBVS data accessible to/through their facilities. Though it will be usable locally, and based on the same metainfo/markup, better local search facilities could be built.

With the new IBVS Data Service we might channel observational data from small observatories to the VO. We think that some of the ideas described here might be applicable to other electronic journals in astronomy.

References

Agosti, D., 2003, Nature, 424, 727

Albrecht, M., Barylak, M., Durand, D., Fernique, P., Micol, A., Ochsenbein, F., Pasian, F., Pirenne, B., Ponz, D., Wenger, M., 1996, `http://vizier.u-strasbg.fr/doc/asu.html`

Good, J.C., Kong, M., Berriman, G.B., 2003, in ASP Conf. Ser., Vol. 295, ADASS XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook (San Francisco: ASP), 89

Holl, A., 1998, in ASP Conf. Ser., Vol. 145, ADASS VII, ed. R. Albrecht, R. N. Hook, & H. A. Bushouse (San Francisco: ASP), 474

Holl, A., 2002, iAstro/IDHA workshop, Strasbourg `http://journals.eecs.qub.ac.uk/tc13/iAstroWebPage/Strasbourg02/Holl.pdf`

Holl, A., 2003, in ASP Conf. Ser., Vol. 295, ADASS XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook (San Francisco: ASP), 321

O'Donnell, M.J., 1993, `http://people.cs.uchicago.edu/~odonnell/Scholar/Technical_papers/Electronic_Journal/description.html`

Schmitz, M., Helou, G., Lague, C., Madore, B., Corwin, G.G.Jr., Dubois, P., Lesteven, S., 1995, Vistas in Astronomy, 39, 272

1.) SOURCE

- Text : † were scanned and OCR-ed, stored in ASCII text files;
  ∗ are in LaTeX source.

- Figures: † were scanned, stored in PS and JPEG;
  ∗ are stored in PS and JPEG.

- Data: stored in simple ASCII text format.

↓

2.) MARKUP & METAINFO

- Figures: † metainfo (including figure keywords) added and stored in auxiliary files together with captions.
  ∗ metainfo is stored in the LaTeX text of the paper (IBVSfigKey)

- Text info: † information from the paper body text (like ephemerides) are extracted to auxiliary files, tables restructured and outfitted with metainfo, stored in text files.
  ∗ marked up with LaTeX macros

- Data: are converted to a standard text format, with metainfo (including data keywords) in the header.

- Biblio: Bibliographic metainfo is available for all issues in ADS tagged format.

↓

3.) IBVS DATA SERVICE - LOW LEVEL

Scripts will serve individual figures, tables, data and other pieces of information (ephemerides) together with metainfo and links to the original paper. Bibliographic metainfo of the paper will be added here. Data and tables will be converted to VOTable format on-the-fly. The input parameter is the IBVS standard filename.

↓

4.) IBVS DATA SERVICE - HIGH LEVEL

Stars are identified with standard GCVS names (if there is one), and a hash table is created with all the files available on a given object. Scripts will serve information on stars, object names will be resolved using the Sesame service of CDS.

Figure 1. Construction of the IBVS Data Service († denotes archive issues, ∗ modern issues).