Information Bulletin on Variable Stars — Rich Content and Novel Services for an Enhanced Publication

Andras Holl
Konkoly Observatory, Budapest, Hungary
holl@konkoly.hu

doi:10.1045/may2012-holl

Abstract

We describe the features of a small enhanced journal, the /Information Bulletin on Variable Stars/ (IBVS). It was founded 50 years ago as a bulletin, changed to a refereed express journal, appeared on the web early on, and is now electronic only, with rich content and services. Technically skilled authors, freely available bibliographic services and discipline-wide standardization — all characteristic of astronomy — form the foundation on which this uniquely enhanced journal is built, enabling it to provide quality services to its research community.

Introduction

The Information Bulletin on Variable Stars (<http://www.konkoly.hu/IBVS/IBVS.html>) was launched as a bulletin in 1961 and distributed free of charge within the small research community of variable star astronomy. Since then it has grown into a small, refereed open access journal. It was a pioneer on the web, appearing online in 1994, but only for distribution of issues in PostScript. Full HTML content was introduced in 2000. The electronic version of the journal has become an enhanced publication, through the gradual addition of research data, a touch of multi-media content, and various services and features. (DRIVER II reported the results of some technology studies in the field of enhanced publications [1 <#1>, 2 <#2>]. Enhanced features of IBVS were developed before these were available, and provide alternatives in certain cases.)

IBVS has published 50-200 articles yearly in the past decades, most of them no longer than 4-5 pages. The articles are published on the web shortly after acceptance, under an issue number. The circulation of the printed journal peaked around 400 issues. The registered users of the electronic version exceed 1,100, about half of them active, and the number of article downloads is 10,000 per month, on average. In the beginning, the paper version was freely distributed to libraries of astronomical institutions and professional researchers, and when it appeared on the web, the electronic version was open access, and a small fee was introduced for the printed version to cover printing and mailing.
costs. The journal appears only electronically now. The launch of the online version coincided, more or less, with the introduction of peer review.

There are three factors that influence our choices of technology. First, IBVS is an express journal. (Stellar variability includes ephemeral phenomena; new findings need to be communicated rapidly.) Second, it is free, it is Open Access, and there are no article processing or page charges. The publisher cannot not afford to maintain a large staff, so everything we do must be done with little work and great speed. The third factor is that it is a very specialized journal published by a research institution, Konkoly Observatory <http://www.konkoly.hu/>, for which the main research profile coincides with the profile of the journal. The authors publishing in IBVS, who are astronomers, could be asked to do typesetting in LaTeX, and the editors, the technical editor, even the typesetting assistant — an undergraduate or PhD student most of the time — are familiar with the scientific field.

Examples of the enriched content and features offered to readers of IBVS can be found in the Appendix <#appendix>.

Reference linking – a different way

Astronomy provides unique possibilities for intelligent literature linking: the availability of BIBCODE [3 <#3>] identifiers and a free bibliographic service, the SAO/NASA Astrophysics Data System <http://adswww.harvard.edu/> (ADS). IBVS uses a linking technology developed in the Centre de Données astronomiques de Strasbourg (CDS). A software developed in-house based on this technology can automatically produce links from the source text of the references list in a paper. Papers are typeset in LaTeX, and references are interpreted, using a complex set of rules and a dictionary of journal name abbreviations. The advantage of the BIBCODE system is that these identifiers can be produced, in most of the cases, from the references used by the authors. Exceptions can be handled by the insertion of the BIBCODE into the LaTeX source by the author or by an editor. Links are produced from arXiv preprint identifiers, and also from DOIs. All literature links point to ADS entries, from where the readers may jump to the full text.

The technique most journals use for linking references requires DOIs, which most of the time cannot be determined from the reference text in the manuscript. Using DOI links would involve manual editorial post-processing of the articles, interaction with CrossRef (CrossRef membership is necessary as well) which involves time, manpower and money that we cannot afford. The technique we use is ideal in our situation, but it would be difficult to implement outside astronomy or physics (where BIBCODEs are not used). However, natural language interpretation of reference texts, like ours, combined with OpenURL [4 <#4>,5 <#5>] technology could achieve a somewhat similar result.

Linking to databases

Besides literature linking, we have links for astronomical databases. The authors (or the editors) might insert special LaTeX macros to produce such links. We provide astronomical object links to the SIMBAD database at CDS, the on-line version of the General Catalog of Variable Stars (GCVS, Moscow), the NASA Extragalactic Database (NED) and dataset (photographic plate) links to the Wide-Field Plate Database (WFPDB) in
Most of these links are "lexical references" to the astronomical objects in question, but WFPDB links are links to actual data used for the article. (WFPDB links will be discussed further in the next section.) SIMBAD, NED and GCVS are all well-known and regularly updated databases in their respective fields. Links do not only help the reader to look up the objects, but help the editors to ensure that object names are not misspelled (astronomical object designations contain a lot of numbers). Furthermore, the arguments of the object name macros — resolvable object names — are reported automatically to the CDS upon publication. In this way we ensure that astronomical object bibliographies are promptly updated with the information in the newly published article.

Research data

Linking peer-reviewed literature to associated data is the core theme of the OpenAIREplus EU project, and a popular topic in the current literature. Collecting associated data files started more than a decade ago at IBVS. We have also collected a few data files from earlier. We encourage authors to submit the data on which the article is based. Making their data available helps to establish the credibility of the article, could be used by the editors or the referee during the peer review process, and might be re-used later on.

If the data is already publicly available somewhere — in a project database, or a data repository — we will link to that. However, we find that it is not the case most of the time. We are willing to store the data and make it available ourselves, for the convenience of the referees and the readers. Because IBVS is a narrow field journal, we think that keeping the research data related to the published articles with the journal is correct. No third party needs to be involved, and the data is at the same location as the article. This arrangement ensures that the data also goes through a peer review process. The editors judging the paper and can view the data, evaluate the quality of the data, and evaluate the meta-data of the data, as well. Both quality control and long-time preservation benefit from the proximity of the articles, the data and the research community — publisher, editors, authors and readers.

IBVS recently adopted an even broader role for long term data archiving. The Variable Star Commission of the International Astronomical Union (IBVS was initiated by this Commission) had an archive for unpublished observations, which was small and sparsely available in digital form. The archive proved too troublesome for larger astronomical data centers. IBVS, together with an amateur organization, the American Association of Variable Star Observers, made an effort to clean up the meta-data and digitize the non-digital documents. The reports of the archive, with lists of archived files, were published in the IBVS, enabling the digitized data files to be linked to these reports.

The data might come as plain text files (with headers containing meta-data), or in the FITS format widely used in astronomy. The latter might contain image or spectral data too.

WFPDB linking provides links to research data (or at least to meta-data of such data). It contains meta-data of photographic plates, and there are also preview images available for the plates linked to from IBVS. Providing full digitized plate content might become practical in the future (a medium-sized astronomical plate digitized with a scientifically meaningful resolution might be 400 MB).
the IBVS, photographic plate identifiers, or observation dates and other meta-data, were extracted from old articles, and now with the help of auxiliary files the readers can check the plate previews and meta-data in WFPDB.

Archiving data and archiving publications are two different tasks. Just as there is diversity in how to describe the different kinds of publications (books, research articles, etc.), there are thousands of different ways to describe data sets, from all fields of science. To display a publication properly you must use only a handful of tools, all easily manageable. Visualizing a complex research data set, especially in the field of natural sciences, is not something an outsider could do. To archive research data, one could opt for using a general data archive, if such archive existed, and the archivists’ task could be simply to check the consistency of the deposited packages. (For example, to check, whether there is a visualization tool that is referred to in the package.) But for long term preservation, general data archives will have a lot of trouble migrating these data and finding new tools. We argue that – at least in natural sciences – only field-specific data archives are sustainable.

Persistent data set identifiers are extremely desirable. The use of DOIs by DataCite is an excellent solution. The only reason we do not use them is that we have no revenue, and could not afford the cost, however small and reasonable they are. But we do have data set identifiers, which are resolvable. One can reach the meta-data and the data set itself by pasting the identifier to the end of an URL.

In astronomy, most archived data can be found in project archives and in astronomy data centers. While the technologies developed by the Virtual Observatory [8 <#8>] project will make accessible data stored outside of data centers. Long term preservation might be better achieved at data centers or journals.

Multi-media content

Multi-media content is not typical of IBVS. It is rare indeed in astronomy. A few IBVS articles do contain animated GIF images, though. Animated GIF does not require additional software, can be viewed with the browser, and there is no demand for anything additional at the moment.

Each figure published in IBVS has its own meta-data, and by clicking on figures readers can access the figure in different formats and check the meta-data. The figure information pages contain links too.

Rich tables

When a paper contains information, such as plots, images, and database links, on a large number of astronomical objects, we arrange this information in HTML tables containing links. One table might contain hundreds of links, however, thousands would be impractical. If that quantity proved necessary once, the search tool could be utilized.

Using third-party visualization tools
IBVS often contains tables with information on a large number of stars in a given field on the sky. Such information is most effectively handled by CIS software (Celestial Information System, as a variation of the Geographical Information System). Instead of using locally developed software, we employ a third-party tool provided by the CDS, the Aladin Sky Atlas [9]. Tables and charts (images) stored at IBVS can be displayed with Aladin by clicking on a link in the HTML version of an IBVS article. The link contains a macro which instructs Aladin where to download the data from, and what to do with it. In the case of IBVS, we use Aladin to display and manipulate JPEG images and tables. The tables are sent to Aladin in a standard XML format developed by the Virtual Observatory community, the VOTable. The reader is able to display objects from astronomical catalogs stored at CDS together with the data coming from IBVS, and interact with the complex dataset.

The "dumb" raster graphics of the IBVS figure can be viewed side-by side with a celestial-referenced image of the same field. Aladin even provides ways to cross-reference the two images.

Aladin is a Java applet, and heavily uses data stored at CDS. Another Java software tool which IBVS could use in the future for a simpler task, the visualization of x-y data plots, could be the VOPlot software developed by the Virtual Observatory India.

The use of third-party tools is a concept rarely found at journals, but well-accepted in the GRID community. It can be regarded as "outsourcing". It can be better to use already available, well-proven software, especially when it comes with extensive services, than developing tools in-house.
Figure 1: IBVS figure (in the left panel) compared with a referenced image of the same patch of the sky using CDS Aladin.

Meta-data availability and visibility

IBVS makes the article meta-data visible to the readers and sends it via email, at the time of publication on the web, to the relevant data centers for indexing, the ADS and CDS. We plan to make the meta-data harvestable by OAI-PMH protocol too; an experimental set-up is already working. The current method is the "push" — we send the meta-data for the bibliographic databases. With OAI-PMH the "pull" method will become available — the databases could harvest our meta-data too. We found that both methods have their merits. With the push method we can send our meta-data to databases not capable of OAI-PMH harvesting. On the other hand, as we regularly maintain our meta-data (correct errors, supply object designations not available at the publishing date, add auxiliary files), and these changes are usually made manually without using the
publishing pipeline, corrections are often not propagated to the databases.

Using OAI-PMH would ensure that all of our meta-data is up-to-date. OAI-PMH would let us communicate meta-data with DOAJ as well. We know that for IBVS, a very specialized scientific journal, the most effective visibility is provided by the astronomical databases CDS, ADS, and GCVS. We frequently get downloads from regular readers browsing the IBVS table of contents. We plan to introduce RSS feeds for the convenience of those users. Adding IBVS meta-data to DOAJ is expected to increase our traffic, or useful visibility, only marginally.

We do not allow search engines to crawl IBVS articles for the same reason. After an initial period we analyzed our access logs and found that search engines generate only "trash" traffic. For non-astronomers, IBVS hits only add to the "noise". A good example is a hit for the search string "movie star" – the IBVS article was most certainly not what the reader was looking for. Experts (professional and amateur astronomers) use ADS and CDS, or follow reference links from articles in other journals to find IBVS. Furthermore, the search engines crawlers add a considerable load to the server, with indiscriminate harvesting. They download all copies of the same article in different file formats to index the full text. We supply astronomical bibliographic databases with meta-data, in a small size e-mail message, where all important information, including keywords and objects, are present in a structured form. The astronomy-specific databases could be instructed to fetch the LaTeX source should they want to provide full text search, thus preventing the unnecessary load caused by the search engines.

We maintain meta-data deeper than the article level. Each figure, and each data file, has its own meta-data, including object information. The readers might search for figures or data files using object names and special keywords.

Advanced search

The availability of rich meta-data makes sophisticated searching possible. Users can search for semantic article components in addition to articles. Another feature of our search tool is the use of name-resolution for author and object names.

It is not unusual for an author to publish using aliases or name variants, and different transliterations from non-Latin alphabets occur. The same astronomical object might have dozens of different catalog designations. Readers can opt for name resolution using local dictionaries or they can use external resolvers. The ADS has author name alias dictionaries, and CDS and the GCVS have object name alias dictionaries that are better maintained than the local ones. If a reader opts for external name resolution, the search string – name – is sent to the external resolver, and all returned variants or aliases are searched for at the IBVS.

The enhanced article as a compound object

As we have demonstrated, IBVS articles contain embedded semantic elements (figures and tables), and reference external elements (data files). The glue between the semantic building blocks are hyperlinks and the file-naming scheme. The component names appear in the readable text of the article, and are present in the components wherever possible
(data files in text format contain the component name, which include the issue number). We do not currently use OAI-ORE [10], but it could be introduced easily.

Information for mash-ups

IBVS provides special figure services intended for creating mash-ups. Other parties can link to these services, embed IBVS figures with concise meta-data and links to their web pages. The GCVS catalog, for instance, often refers to IBVS articles as the source for finding charts (maps) for its entries. It is possible to embed the map in the GCVS web page from the IBVS article instead of merely providing a link. The WEBDA stellar cluster catalog in Vienna contains maps from IBVS too. These can be made dynamic with this feature, however, this feature is not used yet.

Errata

The question of how to deal with errors in published articles is a delicate one. On one hand, articles should remain immutable, but on the other hand, the errors discovered should be clearly indicated. IBVS corrects typos and "technical" problems, indicating the altered nature of the paper, and publishes errata on finding substantial mistakes that are attached to the uncorrected original article.

Marnix van Berchum lists the attributes of enhanced journals [11], with post-publication data among them. We do not have commentaries, nor do we have rankings, but we use the readers as a second line of referees. Errata can be regarded as a form of comments.

Conclusions

Some enhanced features of IBVS extend to the old, scanned issues. The HTML versions of the articles are produced on-the-fly, so newly introduced features of the software also affect the old issues. Active content curation is the attitude at IBVS. This small journal is half a century old, looks like it is produced by an awk script (which is actually the case), but is still young at heart.

The achievements of this journal were made possible by outstanding international collaboration within the astronomy publishing and data community in general, and the Virtual Observatory movement in particular. Small-size specialized journals published by small research units have the disadvantages of scarce financing and lack of many of the resources widely used in the publishing industry, but the closeness to the research community is a definite advantage. Most importantly, field-specific resources are utilized to provide rich services to the readers.

References

[1] Vernooy-Gerritsen, M. (ed.) (2009), "Enhanced Publications: Linking Publications and Research Data in Digital Repositories". Amsterdam Univ.
Appendix

Examples Showing Advanced Services and Features

1 Reference linking:

* See http://www.konkoly.hu/cgi-bin/IBVS?6002 and scroll down to the references section at the end.

Linked text shows the standard bibliographic codes used in astronomy: the BIBCODEs. Journal abbreviations, publication year,
volume and first page can be recognized easily. BIBCODEs can be constructed from the references in the manuscript. Reference links are used not only for the convenience of readers – they are automatically constructed from the manuscript and therefore help the editors who are checking the references. Missing links could indicate a syntactical error or a reference that does not conform to the style of the journal. Links that are not working suggest an error in the bibliographic data, such as the publication year or volume, etc.

2 Linking to databases:

* See http://www.konkoly.hu/cgi-bin/IBVS?6001. Try the SIMBAD and GCVS links at the beginning.

Linked text shows the name of the object. In this case the entry for V963 Persei – a variable star – can be looked up in the SIMBAD database at the CDS, and in the GCVS at the Sternberg Astronomical Institute. Readers can look up objects and get background information. Editors use this feature as well.

3 Research data:

* See the "Data file" link at http://www.konkoly.hu/cgi-bin/IBVS?6001.

The file 6001-t2.txt contains observational data in a plain text file. Lines at the top of the file, beginning with a "#" character contain meta-data. They document the provenance of the data and the equipment used for the measurement. Though IBVS data files do not have DOIs, a consistent naming scheme is used. The filename links this dataset to the IBVS issue No. 6001, and identifies it as Table 2. Some tables also appear in the printed article in a typeset form. If practical, printed tables are duplicated as machine readable data files.

* See http://www.konkoly.hu/cgi-bin/IBVS?1170 and click on the ETABLE links. Click on the KON060 006910 link, and on the PREVIEW image. Enlarging the image one can see the stars. Finding the supernova using the chart in the article is not an easy task!

IBVS No. 1170 reports the discovery of two supernovae – the explosive deaths of two stars. Details of the photographic plates used for the discovery are reproduced in the illustrations of the article. The plates were digitized decades later, and low resolution "preview" images are available at the WFPD. The full digitized images are large, and storage space and network bandwidth limitations do not allow access to the full digital dataset yet. In the future the readers will be able to examine the discovery plates in full detail, including image processing and analysis with software, satisfying the rationale for data publication, which is making the conclusions of the journal article verifiable, and enabling further research.

* See http://www.konkoly.hu/cgi-bin/IBVS datatable?5908-t4, and check the entry for the article in the ADS: http://adsabs.harvard.edu/abs/2009IBVS.5908....1C and then click on the "On-line Data" link. Go back, and go to the on-line article by clicking on "Electronic Refereed Journal Article (HTML)". From the IBVS article, click on the "META-INFORMATION" link at the header to see what is reported to ADS about the data file – the "%f" tag
denotes the data file.

Most readers arrive at IBVS articles following reference links in other electronic journals, or as a result of searching in the ADS or SIMBAD databases. ADS is a bibliographic database. In SIMBAD, users can query what was published on the particular object they are interested in. IBVS is an express journal, and if papers are published rapidly, then researchers should be able to find them rapidly after publication. Meta-data are communicated to ADS and SIMBAD automatically, so a few days after the publication of an IBVS article, users of ADS or SIMBAD will be able to find it. ADS needs to know only the dataset identifier (e.g. 5908-t4). By pasting it to the end of the URL of the IBVS data table service, a link can be constructed. Thus it is possible to resolve the data file identifier to get the file and its meta-data too.

4 Multi-media content:

* See http://www.konkoly.hu/cgi-bin/IBVS?5705 and click on the "Figure: 5705-f6" link and then choose the GIF format.

The figure shows the brightness variation of a star. The variation is not strictly uniform. The course of brightness differs from cycle to cycle. Such complex phenomena can be better visualized with an animated figure.

5 Rich tables:

* See http://www.konkoly.hu/cgi-bin/IBVS?5457, and find the "Table: 5457-t3.tex" link.

This IBVS article publishes a list of stars – about 60 of them. There is a figure (plot of the brightness variation) available for each object, and links are supplied to various databases. Such material can be effectively handled with a HTML table, with links to the figures and the external databases.

6 Using third-party visualization tools:

* See http://www.konkoly.hu/cgi-bin/IBVS?6002. Find the "AladinFindingChart:6002-t1" link, and start the Aladin Java applet. When all data is loaded to Aladin, click on the little red square at the center. The background image comes from Strasbourg. The numbers displayed for the star denoted by the red square come from Budapest.

IBVS data can be visualized using CDS Aladin as a third-party software tool. Data published in the journal is passed to a software tool that resides on the Internet, and can be instructed to display the data using a macro program embedded in the URL in the journal article. The resulting figure is a composite of information coming from the journal, and other data stored at the CDS or elsewhere. Stars in the table of the IBVS article are displayed on the map. Their brightness value, also originating from the IBVS article, could be displayed as well. The reader can access a host of other catalogs as well, and compare IBVS data with information from other sources. By using Aladin, the readers are using a tool more sophisticated than the journal could ever provide, something used widely in astronomy.
* See http://www.konkoly.hu/cgi-bin/IBVS?5901, and click on the first image, then on the "Aladin view of the chart" link.

The "dumb" raster image in the IBVS article can be compared with a referenced image (an image, which has an invisible coordinate grid attached). Charts printed in old IBVS issues are sometimes hard to compare with modern photographic maps. This feature brings CDS Aladin to the aid of the reader to identify the region of the sky on the illustration.

7 Meta-data availability and visibility:

* See http://www.konkoly.hu/pub/ibvs/6001/6001-t2.txt and see the meta-data at the first few lines of the text file. Go to http://www.konkoly.hu/cgi-bin/IBVS?6001 and click on the META-INFORMATION link in the header of the page.

Making the meta-data visible to the readers (including the authors of the article) helps to eliminate errors. Readers can check what was reported about the article to ADS and SIMBAD, or what information is used by the search tool at IBVS. The first example shows the meta-data of a data file, the second that of an article.

8 Advanced search:

* See http://www.konkoly.hu/cgi-bin/ibvs_ast.pl. Choose the "GCVS" option radio button at the top of the page (under Object Name Resolution), and enter "BY Dra" in to the "Object name" field. Press the "Start complex search" button. At the top of the results page you will see the object name aliases returned by GCVS and used in the search. Articles are found in which the "BY Dra" string does not appear, but the star is present under a different name.

Full text search is powerful, but weak at the same time. What we are searching for is an occurrence of a string of characters, not the object or subject they denote. The object contained in the GCVS as "BY Dra" is the same as "HD 234677", the 234677th object in the Henry Draper catalog. When the variability of the star was reported in IBVS No. 122 in 1966, the HD number was used. As a consequence of the article, the object was entered into the GCVS catalog and got the name "BY Dra". Later articles use this name. The IBVS search facility finds the papers about the object if either of the commonly used names are entered into the query field, no matter which of them was used in the article. Search engines can find misspelled occurrences of a word using hash tables with the frequent erroneous versions. IBVS can use local hash tables (alias dictionaries), but can obtain the aliases from other services. GCVS or SIMBAD specialize in keeping aliases up-to-date, so it is more effective to use these external services than to maintain alias lists locally.

9 Information for mash-ups:

* See http://www.konkoly.hu/cgi-bin/IBVSfigure?5921-f2&Format=compact.

The electronic GCVS has links to articles containing finding charts of variable stars. Many links point to IBSV articles. This service of IBVS makes it possible to embed an IBVS figure in a catalog
dynamically. The essential attribution and provenance information is present.

About the Author

Photo of Andras Holl

*Andras Holl* is the IT manager of Konkoly Observatory, Hungarian Academy of Sciences, and curator of the Repository of the Academy's Library at the Library of the HAS. He received his MsC in Astronomy from the Eötvös Loránd University, Budapest. He currently works on the Hungarian National Scientific Bibliography project.

Copyright © 2012 Andras Holl