Innovative or sustainable? Future of the IBVS

András Holl[1],[*], *László* Molnár[2],, *László* Szabados[2],, *Emese* Plachy[2],, *Evelin* Bányai[3],, and *Mihály* Váradi[2]

[1] *Library and Information Centre, Hungarian Academy of Sciences, Budapest, Hungary*
[2] *Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Sciences, Budapest, Hungary*
[3] *Dept. of Physics of Complex Systems, Eötvös Loránd University, Budapest, Hungary*

Abstract. Long-term archiving of scientific information is a well known challenge. We discuss the issue of preservation for journals and their enhanced versions. We use the Information Bulletin on Variable Stars as an example. This express journal, created to satisfy the need of researchers for rapid dissemination of discoveries and results developed to an enhanced journal, with several data journal features. However, the original software developed for the enhanced web version of the journal is not sustainable for the long term. We discuss the possibilities for moving the journal to a standard open source platform, and the trade-offs of the process. Long-term preservation of content, including data, is addressed as well.

1 Introduction

The Information Bulletin on Variable Stars (IBVS) is a small express journal focused on the narrow field of astrophysics, the variable star research. Established as a Bulletin, it became a pioneer on the web, and became an enhanced journal.

The enhanced features of IBVS were innovative, some of them even unique, but, as we had to realize, unsustainable on the long term. As IBVS is a micro journal, without extensive financial resources, and a technical editor doing technological development in his free time, it become impossible to migrate all the existing features to a new platform, when it became necessary. Editors of the IBVS decided to use an OpenAIRE grant for the renewal of the journal, and keeping some key features, abandoning certain enhanced functions, and gaining some new functionalities – like OAI-PMH compliance — the journal will move to Open Journal Systems.

2 Preservation

Preservation of digital or digitized publications - books, journal articles or volumes – needs consideration, attention, resources, but certainly is possible. Institutional or subject repositories, JSTOR, ADS, could provide solutions. In the case of IBVS — as it is a data journal too — preservation of data files should be given consideration too. IBVS contributed to the preservation of the Archives of

---

[*]e-mail: holl.andras@konyvtar.mta.hu

Unpublished Photometric Observations by digitizing some of the material which was available on paper – hand-written, typed or printed by early computers (note, that the data was passed on in printout form, not on magnetic tape). The data IBVS deals with is mostly photometry, which could be well represented in textual form – columns of numbers. These are data files we believe we could preserve for the future. Other data types are less frequently attached to IBVS papers – spectra are sometimes stored in textual tables, images are in PostScript or JPEG – these do not represent a very difficult problem. There are some FITS files – these are standard enough.

What about the complex services (like SIMBAD, Aladin)? IBVS is an enhanced journal, with many fancy features. Is it possible to preserve the Aladin display service IBVS uses for photometric sequences or finding charts? The advanced search features with author or object name expansion using ADS, GCVS or Sesame? These features depend on technologies, and on other parties. They might be re-written using a different programming language, but it requires a lot of effort. SIMBAD, for instance, had many successive generations - CDS was able to migrate it, preserving some important aspects, to new platforms [2]. For IBVS, we had to recognize that we have no means for re-building the enhanced features on a new platform, and might have to loose some of them.

3 History of IBVS

IBVS was conceived during the business meeting of the Commission 27 (Variable Stars) of the International Astronomical Union (IAU) held in Berkeley during the General Assembly of this international organization in 1961. In this meeting, participants accepted László Detre's (then Director of the Konkoly Observatory) suggestion concerning setting-up a new publication for rapid dissemination of information important for variable star community. At that time, researchers of variable stars needed a quicker way of communicating new discoveries, results and announcements than it was possible through the existing journals. The first issue was published under Detre's editorship already in 1961. The IBVS has been edited and published at the Konkoly Observatory ever since. In this 'classical' era, individual issues of the Bulletin were published within a month upon receiving the manuscript.

The IBVS was most popular in the 1970-1980s: about 150-180 issues were published annually. During the almost sixty years of publishing the IBVS, the average production rate was about 110 issues/years.

During the General Assembly of the IAU held in Buenos Aires in 1991, Commission 42 (Close Binary Stars) joined Comm. 27 as sponsoring body of the IBVS. Forming an Editorial Board for the IBVS was also initiated there. Refereeing of the submitted manuscripts by international experts was introduced in 1995.

The electronic edition appeared in 1994, first using the web as a distribution platform of the PostScript issues, then in 2000 it become an enhanced journal. The print edition ceased in 2011, the issue No. 6000 being the last printed one. Since 2011 the journal has been published only electronically.

Following the restructuring of the IAU, the IBVS is now published on behalf of the Commissions G1 and G4 of the International Astronomical Union.

4 Features of the old electronic edition

The enhanced features of the journal were discussed by [1]. Notable are the use of third-party tools and services for visualization of maps and the author and object name searches. The basic idea behind the enhanced version was that the LaTeX source should contain every information needed, the HTML is to be generated from the LaTeX automatically.

The enhanced, HTML version of IBVS were developed with the help of CDS, using technologies developed in CDS.

5 The OpenAIRE grant

OpenAIRE: a FP7-funded initiative to make EU-funded research (e.g., FP7, ERC) open-access, implementing the OA initiative of the European Commission. OA publishing will be a requirement for all H2020-funded projects, therefore OpenAIRE was also aimed at funding the development the underlying infrastructure of open access publishing. One of the projects of OpenAIRE was a call for OA journals that does not collect article processing charges from the authors either. This program provided funding for APC-free journals to implement various technical improvements in their publishing services.

We submitted a successful bid to renew the journal in 2017. Our main driver was the need to replace the outdated website, and to introduce an integrated, on-line editorial workflow instead of the current one. Managing submissions between emails, ftp transfers, and Dropbox directories inevitably led to delays and occasional confusion.

6 Basic features of the new IBVS

As a first step to renew the journal, we registered IBVS at CrossRef and started assigning DOIs to the articles in late 2016, with the help of the Library of the Hungarian Academy of Sciences (MTA). We also renewed our registration with the DOAJ, and will upload article-level metadata to their database.

We selected the new, 3.0 version of the Open Journal Systems software, developed by the Public Knowledge Project, for the new engine of the journal. OJS provides an online submission, review, and editorial workflow system and a publishing platform for us. The public version of the journal will be hosted by the Library of the MTA.

OJS provides many desired features out of the box. These include: OAI-PMH harvestable metadata, management of author ORCIDs, metadata submission to Crossref to register DOIs for the publication, etc.

A requirement of the OpenAIRE bid was the introduction of XML-based article versions that are well suited for machine reading. OJS uses the Lens XML viewer produced by eLife Sciences. This presented us with the greatest challenge: to migrate the archive to OJS and convert the old papers into XML. More than 6000 articles and notes have been published in the journal so far. The first 4100 issues were published in print only and have been later digitized. The rest have been submitted and published electronically too.

7 Technologies used

OJS 3.0 has a plugin for displaying articles in a new innovative way, which makes reading articles online easier. This tool is called eLife Lens which was originally developed for the eLife Sciences webpage.

IBVS uses LaTeX files as source files for its manuscripts. To display the articles with Lens they need to be converted to JATS[1] XML format. There are several LaTeX to XML converters available like LaTeXML, Tralics, LXir, Pandoc, etc. However these were not good for our purposes since we need a converter which converts to JATS XML. Furthermore IBVS has several self defined macros.

---

[1]Journal Article Tag Suite - special set of XML elements and attributes for tagging journal articles

There is a Lua custom writer for Pandoc on github generating JATS XML - but unfortunately there are several issues with it (e.g not recognizing the IBVS macros, not handling some special characters like ; < to its XML/HTML entity value which totally messes up the final XML document).

We decided to write our own converter to produce the proper source files for the Lens viewer. The software is being developed in C. Briefly it parses the document into a Document Object Model (DOM) object first, then writes out the document as a JATS XML file. The converter has 3 parts: the DOM which is independent of the other two parts, the parser and the XML writer both using the DOM part but being independent of each other.

The DOM both reflects the structure of a LaTeX document including the special IBVS macros and partly the expected structure of the XML.

The converter uses the XMLWriter C library to write out the document in XML format which takes care by default about the special characters like an < and raises error if there is a problem with the number of the end tags - if it cannot resolve the problem automatically.

The most complex part - and hardest to develop - is the parser. It uses a series of regular expressions to clean the document from the unnecessary LaTeX macros and commands, convert the special LaTeX letters (greek, accented etc.) to a HTML friendly format, then parse the document in a descending manner. It gets more challenging at the TeX level where it has to take care of the LaTeX markups, inline math mode etc. The program is still under development but it already can partly convert our files (e.g. markup conversion missing, some of the IBVS macros not implemented.) It already has the advantage to convert multiple files at once, parse special IBVS macros, produce a viewable document in Lens, and it will be easy to extend the functionality or write new methods in the future.

Not only the "modern" LaTeX sources are to be transformed to XML, but the plain text source (produced by Optical Character Recognition) of the first 4000 issues too. These texts have neither semantic structure, nor tags. We are developing Python code to collect the metadata of these issues and to mark up the text. The most we could hope for is a semi-automatic process to be developed for the conversion.

The archive issues – either converted from plain text or from LaTeX - need to be imported to the OJS system.

8 Difficulties encountered, results achieved, plans for the future

An unanticipated difficulty in delivering the improvements agreed on in the bid was the bugs we encountered with OJS. Some features, such as issue importing or various plugins did not work as we expected them to, and we had to wait for newer releases to appear with appropriate patches.

The most daunting task in renewing the journal is the conversion and migration of the archive. Given the large number of articles, manual conversion of the files is not feasible. However, automated conversion of the files is not straightforward either. The print and electronic era of the journal present different problems.

The print issues have been digitized and are currently available as scanned images and plain text files. The format of these text files can be wildly different as they closely follow the layouts of the original printed issues. Therefore, without strong cues in the files, even the recognition of the presence of tables or equations can be problematic, let alone automatically formatting them. LaTeX-based files present different problems. Development of the respective conversion algorithms are under way.

9 Conclusion - virtues of simple, widely used technologies

Likewise, the enhanced HTML version of IBVS is a complex machinery heavily dependent on technology. Maintenance and preservation of such services is costly, and as in the present case, might

become unsustainable. We might be able to continue the enhanced HTML version of the journal for a while – surely till the new platform becomes fully operational, and maybe even further, but it cannot be rebuilt with new technology, keeping the full functionality presently available.

We should plan preservation of the information services – databases, journals – we create beforehand. We should designate a core set, and use a simple format, which depends on technology as little as possible. Preservation should not be an afterthought.

As for IBVS, the journal will continue to serve the researcher community, in a somewhat different form. We hope that the new features and design will please the readers, and that eventually many extra services could be rebuilt using different technology - as OJS plugins.

References

[1] Holl A., D-Lib Magazine **18**, 4/5, 4 (2012)

[2] Wenger, M. et al., *ASPC 351*. (ASP, San Francisco, 2006) 662