Manuscript submitted for the proceedings of the 5th Belgrade International Open Access Conference, National Library of Serbia, May 18, 2012 – May 19, 2012

Enhanced journals - a case study with general remarks

András Holl

holl@konkoly.hu
Library of the Hungarian Academy of Sciences, Arany J. Str. 1., Budapest (Hungary)
Konkoly Observatory, Hungarian Academy of Sciences

Abstract

The Information Bulletin on Variable Stars - a small astronomy journal published in Hungary - was an early Open Access journal. Now it offers several enhanced features to its reader community. Relying on the rather unique publishing environment existing in the field of astronomy, and on software developed locally, this journal is markedly different from other enhanced journals in certain aspects.

We explore the key features of enhanced and common electronic journals: reference linking, database connections, data linking, multi-media content, feedback from the reader community, quality control. We argue that while exploring new avenues of scientific publishing, one should conservatively preserve some traditional values and features.

Some aspects of article disassembly - dealing with items smaller than the usual basic publication unit, the article - are explored too. Figures, for example, are article components which might be re-used, used outside the original context.

Introduction

The Information Bulletin on Variable Stars (IBVS) is a small electronic journal published by Konkoly Observatory, Hungarian Academy of Sciences. It covers a narrow research field in astronomy: the study of stellar variability. Originally it was started as a bulletin in 1961, it appeared on the web in 1994, and became a refereed journal around that time. The paper version was discontinued in 2011, since then IBVS has been an electronic-only journal (http://www.konkoly.hu/IBVS/IBVS.html).

Enhanced journals & IBVS features

Features that enhanced publications offer usually include the following: i.) extensive linking of external resources; ii.) availability of research data; iii.) multimedia content; iv.) post-publication material (see the information on the SURF Foundations EJME project).

Reference linking

Reference linking is not considered as an enhanced feature. Most electronic journals offer linked references - the usual method is via CrossRef. IBVS does have reference links, but relies on a different support infrastructure: the bibliographic identifier standard in astronomy (Bibcode, see Schmitz et al., 1995) and the NASA/Smithsonian Astrophysics Data System (ADS, see Kurtz et al., 2000), a free bibliographic database. Bibcode identifiers are created from the source text of the article using technology initially developed at the Centre de Données astronomiques de Strasbourg (CDS), and tailored to the needs of IBVS. The rationale behind technologies used at IBVS is discussed in more detail by Holl (2012a). Reference linking is an essential feature. The IBVS method is probably not adaptable generally - it is beneficial for an express journal, but might not be suitable elsewhere.

Object linking

IBVS provides links to several astronomical object databases and electronically available catalogs, including CDS Simbad (Genova, 2007), the NASA Extragalactic Database and the General Catalogue of Variable Stars (Moscow). In this case the objects need to be tagged by the authors. Tagged objects are automatically reported to CDS and ADS (both of these astronomical databases link back to IBVS). All object links are generated from standardized object names. Object linking is a desirable feature for electronic journals - but it assumes the existence of object databases, lexical resources, preferably freely available.

Research data

Creating and maintaining data repositories is a task much more difficult than in the case of publications. Data is best archived by specialists of the given field. IBVS decided to store and make available the data used for the articles. Each data file has its own meta-data, and an unique identifier. (IBVS does not assign DOIs yet, neither to the articles, nor to the data - we can not afford even the minimal cost and manpower increase it involves.) Small volumes of research data are usually present in the research articles - either in the text, or in tables or figures. These data are prepared for human consumption. IBVS tries to put a machine readable data file behind tables and figures, wherever it is applicable.

DOIs offered by DataCite could be used by other enhanced journals, if the data is made available in a repository. If a journal chooses to store research data related to the articles, some database functions need to be implemented, and questions about the long-term preservation of data need to be addressed.

Research data for IBVS is most of the time simple, could be easily stored in plain text files.

But more complex data structures present challenges. Journals might understandably defer from the task of archiving such data. Packaging would most probably be part of the solution, and external, dedicated data archives could be preferable for complex data.

Data linking

The Wide Field Plate Database (WFPDB, Sofia) contains meta-information on astronomical photographic plates stored at several plate archives around the world. For certain plates low resolution preview images are presented, and occasionally the full digitized plate is available too. IBVS offers links to WFPDB entries, in cases we could identify the serial numbers of plates used for a given article (Tsvetkov et al., 1997).

There are not many data repositories yet offering free access. Here the use of standard, unique data identifiers (DOI or other), and standard data formats should be emphasized.

Multimedia content

Multimedia is not widely used in astronomy publications. However, IBVS does have some animated GIF figures. Enhanced journals should take care of using standard file formats to ensure platform-independence, easy access and the possibility of preservation. Wide availability of software for the given format is also of great importance. Animated GIFs can be viewed with any web browser in this particular case.

Post-publication material

The only type of post-publication material IBVS offers are the errata. All the substantial errors noticed after publication result in an erratum, attached to the original article. It is possible for an enhanced journal to provide a forum for commentaries, but our view is that such comments have only a limited audience - those few researchers who visit the journal website regularly. We think that most readers access an article following a link from the references section of another article, or from a bibliographic database, and they often choose the PDF, thus commentaries mostly present in the HTML are not visible for them. The real forum for comments is the whole scientific literature: if an article needs serious commentaries, reflections, they should appear in other articles. On the other hand, while it is important to facilitate scientific discussions, the content of these is different from the literature. Now it is possible to preserve and make available the content of mailing lists, our view is that what matters, should appear in the mainstream literature, and some exchanges of opinions - while these play an important scholarly role - should be left to oblivion. References to "private communication" or mailing lists should be the exception, not the rule.

Use of a third-party data visualization tool

IBVS articles often contain long lists of objects with coordinates and other characteristics which could only be visualized using a spatial information system. As such a system exists in astronomy - the CDS Aladin Sky Map (Boch, 2011) - IBVS relies on that. Charts and star lists are passed to Aladin to enable the reader to interact with the data. In our view there is a need for such GRID-like tools in other areas too. IBVS tables are transferred to Aladin in a standard data format used in astronomy: the VOTable (Ochsenbein et al., 2004)

Meta-data

Providing meta-data for the readers and for the bibliographic services is a must for every electronic (or indeed, printed) journal. IBVS presently supplies meta-data for each issue in the tagged format of ADS. We are ready to provide DublinCore meta-data through OAI-PMH too.

Abstracts

Abstracts form a special part of meta-data. At IBVS, abstracts were introduced for the sake of ADS (and other bibliographic services). These abstracts are the usual, textual summaries of the paper. Some journals recently introduced visual abstracts - figures, which catch the eye of the reader and might tell more than a paragraph of text. Both kinds of abstracts are for humans. One might ask whether journals could provide abstracts for machines? Abstracts which contain the summary, the most important conclusions of the article formatted in a formal language. IBVS regularly publishes short discovery notes on newly found variables. Such information should be digestible for machines too. In a semantic web scenario reports on local cricket or football matches should contain a computer readable abstract on the scores. Scientific journal articles should be linked to the data web too, distilling the most important findings to machine readable format.

Quality control

Alternatives for the current peer review system are being considered. Reader voting - one of the proposed alternatives - does not provide means for improvement of the paper. Open review might have its advantages, but we think it does not relieve the editors of certain tasks: the competence and the independence of the referees must be established anyway. IBVS employs the traditional refereeing system (the single-blind version). However, IBVS benefits

from being an OA journal, with a rapid publication cycle: errors caught by the readers are corrected in the Errata attached to the papers.

Enhanced journals present further problems from the aspect of peer review. How could referees examine all auxiliary materials? Providing the necessary tools should be the lesser problem - some tools at least should be provided for the readers anyway. Other tools - which help the systematic checkout of the material, and not present in the reader's toolkit might be more difficult to provide. IBVS has some tools provided by CDS for object name checking and some other tools developed in-house for discovery note validation and time series manipulation, but referees could not use these. Peer review takes up precious time of the referee, and the whole process delays the publication of the article. Will the referee take the time to go through all the - possibly extensive - auxiliary material? Maybe this is the point where readers could be employed, after publication, as a second line of referees. At IBVS the referee usually sees the PDF of the article, only editors do scrutinize all auxiliary material.

Long term archiving

Enhanced features of the electronic journals depend on many external factors. Longevity of database links or the sustainability of external software are hard to guarantee for the long term. We think that preserving the enhanced features of the journals is a great challenge. Long term preservation is best served by following a proactive approach. Articles in enhanced journals contain informational elements (new data, new findings) and presentational elements (like multimedia), and the two are not easy to separate. Nevertheless, the former is more important, and easier to preserve than the latter. Traditional journals are easier to preserve than enhanced ones, and maybe it is enough to preserve the core components (and the data).

IBVS articles still have - and will have - a self-consistent core, the PDF article, which are preservable on the long run. PDF was chosen because it is analogous to the printed article. We plan to use PDF/A, and we are seeking for a suitable dark archive within the astronomical community - the obvious choice is the ADS. For the data CDS could be a suitable place for archiving. Links in the HTML version, or the Aladin visualization of IBVS data are not necessarily preservable.

Article disassembly, compound objects, aggregation

Scientific journal articles can be disassembled to smaller components - semantic building blocks - and these components might be reassembled to provide a different view of the article, or used independently (see Holl, 2012b). Certain components of IBVS articles could be accessed and used independently of the whole article, e.g. figures, tables and meta-data. All figures and some tables (data files are regarded as auxiliary tables) have their own unique

identifiers and are accessible individually using these identifiers. Each figure and some tables have their own meta-data, including object names. We prefer to have the data available in tabular form for the relevant figures.

Figures are those article components which could be re-used, used out of the context of the original article. With proper attribution and reference to the origin, figures should be reproducible. While using CC licenses for scientific journal articles does not seem appropriate, it should be considered for figures.

Relations between article components and auxiliary material could be mapped using OAI-ORE. IBVS does not use this technology yet. Another possibility of creating compound objects from the article and its auxiliary material is by employing some kind of packaging technique. These are discussed in Enhanced Publications (Vernooy-Gerritsen, 2009).

Advanced search

There are two features that make IBVS search unusual. One of them is the ability to search and retrieve data files and figures. The other is that we have implemented a system which enables searching for object and author name aliases. Aliases could be fetched from external dictionaries.

Usefulness and cost of the enhanced features

Though we have not investigated the reception of the enhanced features thoroughly, it is possible to compare the impact of the journal with other journals in the same field. Such comparisons are favorable for IBVS, and access statistics showed an increase in the downloads with the introduction of the HTML version. However, it is hard to disentangle the effect of the enhanced features. Enhanced journals are still in their infancy, and probably only time will tell which features are really important. In our case the enhanced features were developed by only a few people: the technical editor and some students, using only a fraction of their time. To some extent the development was an exercise which enabled the developers to learn about information technology, services and standards in the field of astronomy. Larger journals, future enhanced journal projects might follow a more rigorous development process.

Plans for improvements

There are several improvements we intend to make for IBVS. OAI-PMH meta-data distribution would help bibliographic databases, while RSS feed would be useful for readers.

We plan to implement Open Researcher and Contributor ID-s as soon as possible. The advanced search tool would need a name preprocessing filter, and for short notes we plan to create a web submission tool.

Conclusions

Features offered by IBVS rely on the special environment of astronomy: field-specific standards & file formats are used and open databases provide linking opportunities.

Enhancements of scientific journals are presently offering “extras”, which are interesting, but mostly experimental features. A number of questions should be answered before these could become main-line accessories of the articles. The other difficulty enhanced journals face might be the lack of free resources: databases in the given field or data repositories. In this regard astronomy is in a very fortunate position. Experimenting with enhanced journals is important nevertheless.

References

Boch, T.; Oberto, A.; Fernique, P.; Bonnarel, F. (2011). "Aladin: An Open Source All-Sky Browser". *ASP Conference Proceedings*, 442, 683

Genova, F. (2007). "The Many Faces of SIMBAD", *ASP Conference Series*, 376, 145

Holl, A. (2012a). "Information Bulletin on Variable Stars — Rich Content and Novel Services for an Enhanced Publication". *D-Lib Magazine, 18, 5-6* (in press)

Holl, A. (2012b). “Article disassembly - new ways to handle information in publications”. *Serdica Journal of Computing* (accepted for publication)

Kurtz, M.J., Eichhorn, G., Accomazzi, A. et al. (2000). "The NASA Astrophysics Data System: Overview". Astronomy and Astrophysics Supplement 143, 41. DOI:10.1051/aas:2000170

Ochsenbein, F., Williams, R., Davenhall, C. et al. (2004). “VOTable: Tabular Data for the Virtual Observatory”. *ESO Astrophysics Symposia*, Toward an International Virtual Observatory, p. 118. DOI:10.1007/10857598_18

Schmitz, M., Helou, G., Dubois, P. et al. (1995). "NED and SIMBAD Conventions for Bibliographic Reference Coding". In D. Egret and M.A. Albrecht (Eds.): Information & On-Line Data in Astronomy. Kluwer.

SURF Foundation, Enhanced Journals...Made Easy! http://www.surffoundation.nl/en/projecten/Pages/EJME.aspx

Tsvetkov, M.K.; Stavrev, K.Y.; Tsvetkova, K.P. et al. (1997). “The Wide-Field Plate Database : Present Status and Future Development”. *Baltic Astronomy*, 6, 271

Vernooy-Gerritsen, M. (ed.) (2009). Enhanced Publications. Amsterdam University Press. DOI: 10.5117/9789089641885