

## **Urania, a Linked, Distributed Resource for Astronomy**

Peter B. Boyce

*American Astronomical Society, Washington, DC 20009, USA*

**Abstract.** Above all, the electronic information environment is inter-linked. When done effectively, links knit together the scholarly references, citations and data sources for the user as never before, bringing the whole web of distributed information resources right to the reader's desktop. A certain amount of infrastructure and degree of cooperation is needed to make this happen effectively. The field of astronomy and astrophysics now benefits from the development of such an infrastructure, which has been named Urania. Although Urania has been working effectively for three years, astronomy's growing needs for greater breadth of coverage will force changes to be made in the Urania infrastructure.

### **1. Introduction**

The electronic, interlinked world of scientific publishing is very different from the world we have known of individual journals containing individual articles. The differences have major implications for the traditional journals, the traditional database providers and the astronomy libraries. Preparation, distribution and maintenance of electronic documents requires an interdependence of all steps in the information chain – from author to reader – which is unheard of for paper journals. Small changes in the electronic manuscript may have enormous consequences later in the production process, introducing complexities which can, and should, be avoided by establishing effective feedback and communication among all parties.

The need for effective communication up and down the information chain suggests that we can function best if we adopt a new mode of working. Instead of dealing with vendors and customers in a confrontational mode, we, at the American Astronomical Society (AAS), have found that the best results are obtained by viewing electronic publishing as a collaborative and cooperative venture.

### **2. The Electronic Environment**

Five important traits characterize today's environment for electronic information dissemination. In general the situation is markedly different from the milieu of traditional paper journals.

1. The electronic information resources are widely distributed, yet are heavily interlinked, and will become even more so. Our readers rate links to

references and citations in the AAS journals as the most important feature. As publishers begin to add electronic value to their journals through links to additional resources, the ability to link will become an even more integral part of any electronic publication.

2. Some features of the paper journals must be retained. Scholarly integrity is one such feature. With the growing amount of information of questionable accuracy appearing on the Web, and access to information becoming available to readers who may not have the capability to judge for themselves the reliability of such information, the scholarly electronic journals will become welcome islands of quality in the growing sea of information available to readers.
3. The electronic environment is changing relentlessly at a breakneck pace. Driven ever onward by advances and innovations in technology, the electronic journals will have to adopt changes and innovations at an unprecedented pace. As the “computer generation” readers begin to dominate the audience, they will expect the scholarly journals to remain close to the edge of advancing technology.
4. Readers will also come to expect that the information on the Web represents the latest, updated material. Yet, the scholarly journal implies an integrity of the material – that it has remained unaltered since it was accepted for publication. These two conflicting requirements do not seem to be compatible. We can expect to see a new type of information resource arise in astronomy, similar to the genome and protein molecule databases in biology, which are a compilation of up-to-date, and presumably best, estimates of the values of measured information. Such new online sources of the “latest and best measurements” will exist side by side with the “traditional” scholarly electronic journals, which remain unchanged after publication. But even the traditional scholarly journals will have to point to the latest citations and updated material in their electronic versions.
5. Finally, the preservation and archival maintenance of electronic material will become more and more of a problem. Preserving effective access to the individual articles as the technology changes will be difficult. But, as electronic journals incorporate more and more links, preserving the links as well as the textual material will be even more difficult. Yet, this will have to be done. In the electronic era, it will be no longer sufficient to find a storage location for a file or two where the material is stored and left like a book on a shelf. Electronic material will have to be actively managed to prevent deterioration of the material to the point where it becomes unreadable (Garrett & Waters 1996). This situation is not unfamiliar to librarians who have witnessed the physical deterioration of journals from the 1920s which were printed on acidic paper and are now crumbling to dust under the reader’s fingers. The same thing happens in the electronic era, only faster. No storage medium lasts forever. The electronic journal has the added complexity of maintaining working links. This is a serious problem to which there has not been enough consideration given.

We are working in a new environment, and the implications how we must change to make effective use of the new capabilities are enormous. Electronic documents will be assembled from pieces located in several places. Even today, one year of the *Astrophysical Journal* is made up of about 250,000 interlinked files. We will soon see electronic documents assembled on the fly, to match the requirements of the reader's capability to access and process electronic information. Different readers could see a technically different presentation.

There will be a wide, and growing, variety of electronic features in the journal of the future. Doctors will see video demonstrations of heart operations. Ornithologists will hear bird songs. Readers of clinical medical journals will be able to calculate risk factors for individual patients by going to online journal articles and plugging information into "live" equations. Finally, astronomers will be able to manipulate 3-D images of star clusters, and replay simulations of galaxy collisions contained in refereed journal articles.

All these things have been demonstrated today in single articles. The future will see such things become commonplace in the journals. We are not used to dealing with journals in these forms. All of us, authors, publishers, librarians and database providers will have to adapt a new mind set if we are to deal effectively with the changing nature of information transfer in the electronic age. It is very difficult for us to envision the enormity of the long term changes which are happening today. Historically, we have always overstated what will happen in the short term as a result of the introduction of new technology. But, we have consistently failed to envision the effects of the fundamental revolution brought on by the growth of the World Wide Web.

This LISA III conference is but one example of how the Web has affected how we do our business. The conference was put together by individuals located in Munich, Washington, Baltimore, Charlottesville and Tenerife, all without using paper mails. Hotel reservations, conference registration, and distribution of the program were all accomplished over the Web. This is a remarkable change over the last decade. The same change is happening in the scholarly information enterprise, and we in the scholarly publishing enterprise have to abandon many of our past habits and older modes of thinking if we are to remain relevant and successful.

### 3. Today's Electronic Journals

Astronomy's electronic information resource has three parts, the electronic journals, the Astrophysics Data System (ADS, <http://adswww.harvard.edu/>), and the various astronomical Data Centers such as the Centre de Données astronomiques de Strasbourg (CDS, <http://cdsweb.u-strasbg.fr/CDS.html>) and the NASA Extragalactic Data Center (NED, <http://nedwww.ipac.caltech.edu/>). The ADS – with searchable abstracts and full text of the historical papers – is described in greater detail elsewhere in this volume (Eichhorn et al. 1998).

Let us first consider the journals. The AAS now has been publishing an electronic version of the *Letters* section of the *Astrophysical Journal* (*EApJL*) since 1995. Boyce & Dalterio (1996) and Boyce et al. (1997) have described the philosophy behind the various features of the AAS journals. Significant effort has gone into making the journal useful for the readers. We use no unnecessary

graphics. Excessive graphics may look “cool”, but they slow down the transmission of useful information. There is an HTML version for browsing and a PDF version for printing out page images. The HTML version of the short articles in the *EApJL* has proven to be about five times as popular as the PDF version. Apparently readers use the HTML version to browse on the screen and the PDF to print out the articles of greatest interest to them.

The readers are interested in seeing the material as rapidly as possible. To this end, the AAS and the University of Chicago Press have worked to revise the whole production process to focus on producing first the electronic version, from which can be derived the paper version which requires a longer time scale for publication. The electronic version of the papers are posted article by article as soon as they are ready – within three weeks of acceptance by the scientific editor. This is a major breakthrough in the dissemination of information.

But, above all, the journals published by the University of Chicago Press are characterized by an abundance of links. These journals include the *Astrophysical Journal* (including the *Supplement Series*, <http://journals.uchicago.edu/ApJ>), the *Astronomical Journal* (<http://journals.uchicago.edu/AJ>) and *Publications of the Astronomical Society of the Pacific* (<http://journals.uchicago.edu/PASP>).

#### 4. The Value of Links

Links are crucial in an electronic journal for two main reasons; for convenience of navigation and reading within an article, and for linking to resources outside of the article. Navigational links are important for readers who are skimming an article on the screen. On-screen reading is generally considered to be more difficult than reading a conventional journal. Therefore, an electronic journal should help the reader to extract useful information as easily as possible, starting with a table of contents for each article with links to the sections, and back to the table of contents. Such two-way links are particularly helpful to electronic readers, but are not uniformly provided by many publishers.

In general, editors and publishers seem to cling stubbornly to stylistic criteria which were laid down decades ago, but are not nearly as effective in the electronic environment. Providing links to footnotes, references, figures and tables are essential, but it is even more useful to provide links back into the text from each of these sections. Two-way links provide new ways for a reader to assess if the whole article is worth reading. They are browsing aids which can only be provided in electronic versions of an article. For example, one way to browse an article is to download the references, look to see if the author has referenced the reader’s previous work, and, if so, have the capability to jump into the text where the author discusses the reference. The browser’s “back” button is of no use whatsoever in such a case.

Other readers may want to look first at an article’s figures – which thus should be able to be downloaded separately without waiting for the whole article to download – and similarly jump into the text for a further explanation of the figure. Such new methods of skimming an article’s contents and assessing quickly its relevance to the reader can only be done in the electronic environment, and only if the publisher provides the necessary formatting and links which make this possible.

As useful as internal links are, those links to resources outside the article are judged by reader feedback to be even more important. Links to online abstracts of the referenced articles are a real boon to readers. Links to machine-readable data tables are particularly important for astronomy. And, as time passes, links to future articles which cite the article should be added. The added value of an up to date citation list will grow with time.

Abstracts of references and citations are wonderful, but the next logical step is to be able to access the full text of referenced articles, a simple matter to accomplish directly to articles in the same journal, and only slightly more difficult when linking to articles in journals produced by the same publisher. As this is written in 1998, three years after the online *Astrophysical Journal Letters* began to offer links to referenced articles, the larger commercial publishers are beginning to assemble collections of their journals for this purpose. However, despite the promises and advertising hype, the promise of effective linking held out by such publisher collections has, in most cases, yet to be actually achieved in an effective way. Astronomy is one field where this has been accomplished for about half the important peer-reviewed journals.

But, readers do not want to limit themselves to the offerings of just one publisher. They want to be able to link to all the relevant literature at the click of a mouse button. Providing links to the works of more than one publisher is a more difficult job, requiring a level of cooperation which is not normally found between publishers. Inserting links one at a time by hand is prohibitively expensive. The only solution is to insert links during the publishing process using a high degree of automation. The key to doing this effectively is to adopt common, open standards for naming articles, to provide a name resolution service, and to work through an abstract service which can serve as a centralized location for resolving links to the material from various publishers.

## 5. Urania

Working with the NASA-supported ADS and the global set of astronomical data centers, we have developed a working system of common standards, naming conventions, and cooperative protocols which have made it possible to link the astronomical literature together into a system of interlinked resources which bring information to the researcher's desktop anywhere in the world where there is an effective connection to the Internet.

To emphasize the importance of this cooperative effort, we have called this enabling infrastructure "Urania," named for the muse of astronomy. Urania is not a collection of objects. It is not a product to which libraries can subscribe. It is not a consortium. It is the underlying infrastructure which makes possible the interconnectivity which characterizes the electronic dissemination in astronomy.

Urania is now based on a "Bibcode" identifier which is derived from a volume, page year naming scheme and is ideally suited to identifying articles published in a scholarly journal. The bibcode, also known as "Refcode" was developed during the 1980s as a cooperative effort among the astronomical data centers, notably the NASA Extragalactic Database (NED). A bibcode can be interpreted by a human reader, which was important during the early days, but is much less important now that a significant fraction of the astronomical

literature is available online. In operation, the Urania participants make up their own bibcodes following the open, standard rules which everyone has agreed to. This has been successful until now because the significant astronomical literature is contained in a limited number of journals, and the number of publishers and data centers involved is small. The bibcode system has served astronomy well, and it provided a simple, workable method for us to get started in developing methods for inserting effective links into electronic documents. However, as we strive to include more of the literature, specifically, monographs and conference proceedings, we find the bibcode system does not serve as well. We will have to move to another system in the near future. In the meantime, however, the science of astronomy is well set up to reap the benefits of a distributed, interoperable system for the exchange of information.

The ADS has a system of searchable abstracts for astronomy's core literature which has proven to be a popular resource (Eichhorn 1998). The data bases provide a system by which information about a specific object can be retrieved, along with links to the articles where the specific data were published. The ADS is well along in a project to scan the full text pages of the major historical literature and make them available online, with the accompanying reference and citation lists which are linked to the articles as well as to the current electronic journals (Kurtz & Eichhorn 1998). And, finally, the current electronic journals are linked to each other and to the ADS system of abstracts and full text page images, and the holdings of the data centers.

The smoothly functioning links which connect the references, citations, data and historical literature are the result of a standardized naming system, excellent cooperation among the various organizations which are providing the information, and a name resolution system by which a link can be directed to the symbolic name of an article without having to hard-code the actual physical location of the article (or piece of data). The symbolic names also make it possible to maintain mirror sites for all the major online information resources which the Urania coalition provides. At this time mirror sites in the U.S., Europe, and Japan speed up the response time.

## 6. Other Identifiers

Another system which fulfills the same purpose for medical literature as Urania does for the astronomical literature is based upon the Medline database operated by the U.S. National Library of Medicine (<http://www.nlm.nih.gov/>). In this case, Medline assigns an identifier to each article. Unlike the bibcode, which is generated according to a known set of rules, there is no way to know, *a priori*, the Medline identifier for a given article. In order to insert a link to Medline, a publisher queries the Medline database to find the identifier. In the tightly coupled world of today's Internet, this system works well, and has the advantage of not requiring that anyone wanting to make a link to an article remember the sometimes arcane rules by which the bibcode is generated.

Another standard naming scheme which is attracting attention is the Digital Object Identifier (DOI, <http://www.doi.org/>). Originally launched by the American Association of Publishers, this is a broadly based naming scheme in which each publisher gets an identification number from the DOI organization

and subsequently becomes the naming authority for all the items they produce. As currently structured, the item, or article, names will reside in one centralized database which seems unduly cumbersome in the present environment of distributed, WWW-based information resources. Moreover, the DOI is designed to track usage and collect fees, not to enable the building of links. However, with the weight of the mass market publishers behind it, the universal adoption of some form of the DOI appears inevitable. But, in order to be useful to the scholarly community, for which links are all important, the DOI will have to provide an automated query system by which publishers will be able to build links to references and citations. In the meantime, astronomy can continue to use what we have. One of the fundamental concepts of the Urania infrastructure is the use of name resolvers. The actual name we use to identify articles is of no particular importance since we can resolve multiple names for the same article. Consequently, astronomy is in a good position to be able to adopt any reasonable standard which will eventually emerge.

## 7. The Future Evolution of Urania

What is the future for the Urania Collaboration? As more journals become available online from other publishers, we will have to begin to interact gracefully with additional journals, particularly with those outside of astronomy. Rather than teaching the whole world how to calculate a bibcode, we will have to develop systems by which other publishers can query us to receive a standard name for our articles by which they can be linked.

Moreover, it is clear from current discussions on the bibcode mailing list that calculating bibcode according to a set of rules is no longer viable. The Urania coalition will have to begin to generate standard names for digital objects (chunks of digital information, such as articles), and to supply them through an automated query system such as Medline now does. This implies that Urania will have to develop a naming authority system for astronomy, and begin to assign names which may no longer be humanly readable. We can keep the calculable and human readable bibcode for journal articles, but outside of the serial literature, arbitrary names will have to be assigned by one central authority.

We also have to work with other disciplines to integrate our naming authority into systems used by other disciplines. Maybe we will adopt the DOI. Maybe we should use Medline to assign identifiers. Maybe we should set up our own system which will cooperate with other discipline-specific naming authority systems. Whichever path we choose, it is clear that astronomy can no longer remain as independent as it has over the last few years. But, with three years of use based upon an open and flexible set of standards, the Urania Collaboration is set up to evolve gracefully and keep up with changing needs.

**Acknowledgments.** Without the cooperation and contributions of the Urania coalition, astronomy would not have the effective system it now has. Without the cooperation and encouragement of the editors of the AAS journals, we would not have the high quality electronic journals which are a critical element of the Urania collaboration. Sarah Stevens-Rayburn of STScI and the astronomical library community have provided helpful insight into the needs of the users. And, most of all, I am indebted to Evan Owens of the University

of Chicago Press and Chris Biemesderfer of ferberts associates who have been valuable members of the AAS Electronic Publication Development Team and have provided the clarity of vision and the technical know-how to actually make this all happen.

### **References**

- Boyce, P. B. & Dalterio, H., 1996, *Physics Today*, 49, 42,  
<http://www.aas.org/~pboyce/epubs/pt-art.htm>
- Boyce, P. B., Owens, E. & Biemesderfer, C., 1997, *Serials Review*, 23, 1,
- Eichhorn, G., Accomazzi, A., Kurtz, M. J. & Grant, C. S., 1998, *The Astrophysics Data System*, these proceedings, 277
- Eichhorn, G., 1998, *Linking in the Astrophysics Data System*, these proceedings, 118
- Garrett, J. & Waters, D., 1996, *Preserving Digital Information: Final Report and Recommendations*, <http://www.rlg.org/ArchTF/>
- Kurtz, M. J. & Eichhorn, G., 1998, *The Historical Literature of Astronomy*, via ADS, these proceedings, 293