

Electronic Publication and Data Distribution for the Five College Astronomy Department

Karen M. Strom

Five College Astronomy Department,
University of Massachusetts,
Amherst, MA 01003, USA
kstrom@hanksville.phast.umass.edu

Abstract

The Star Formation Group at FCAD has begun to make use of the World Wide Web to explore the advantages of hypermedia presentations for the distribution of preprints, Ph. D. theses and observatory publications.

As a byproduct of this work, a method for displaying subscripts (*e.g.* M_{\odot} , $\Delta\nu_D$) and superscripts (*e.g.* T^4 , $C^{18}O$) has been developed.

This set of bitmaps is publically available.

1 Introduction

Most of the professional journals in astronomy are now accepting submissions in \LaTeX . It is clear that advances in manuscript preparation have not been matched by increases in distribution speed — the first and most easily achievable gain which should result from some standardization of the format for manuscript submission.

There are enormous advantages to having papers available electronically with *their associated data sets*. They can be made available almost instantly upon the completion of the refereeing process. They are easily searchable, and, if the tabular material is in an HTML document (if small) or other easily interpretable form, the data is instantly accessible. Large data sets can be made available in standard, generally accepted formats. More sophisticated graphics are easily and cheaply included. Links can be built into the paper to references and abstracts that are on-line as well as to other non-standard materials not usually available to readers of the papers.

2 Preprints and Ph.D. Theses

Since astronomy journals are now accepting manuscripts in \LaTeX , the obvious method for conversion of these manuscripts into HTML is by use of Nikos Drakos' [1] \LaTeX2HTML . This *perl* script will translate a \LaTeX manuscript into a set of linked HTML pages, translating the mathematical expressions into transparent GIFs, placing these images in the text as in-line images. However \LaTeX2HTML may create many copies of the same image when the same character is reused in the document, particularly for commonly used symbols and expressions of units, such as cm^2 or s^{-1} . Since all images were installed with `ALIGN=BOTTOM`, *e.g.* cm_2 or s_{-1} , the resulting text could be difficult to read. It is also true that math mode expressions including subscripts would appear to float above the rest of

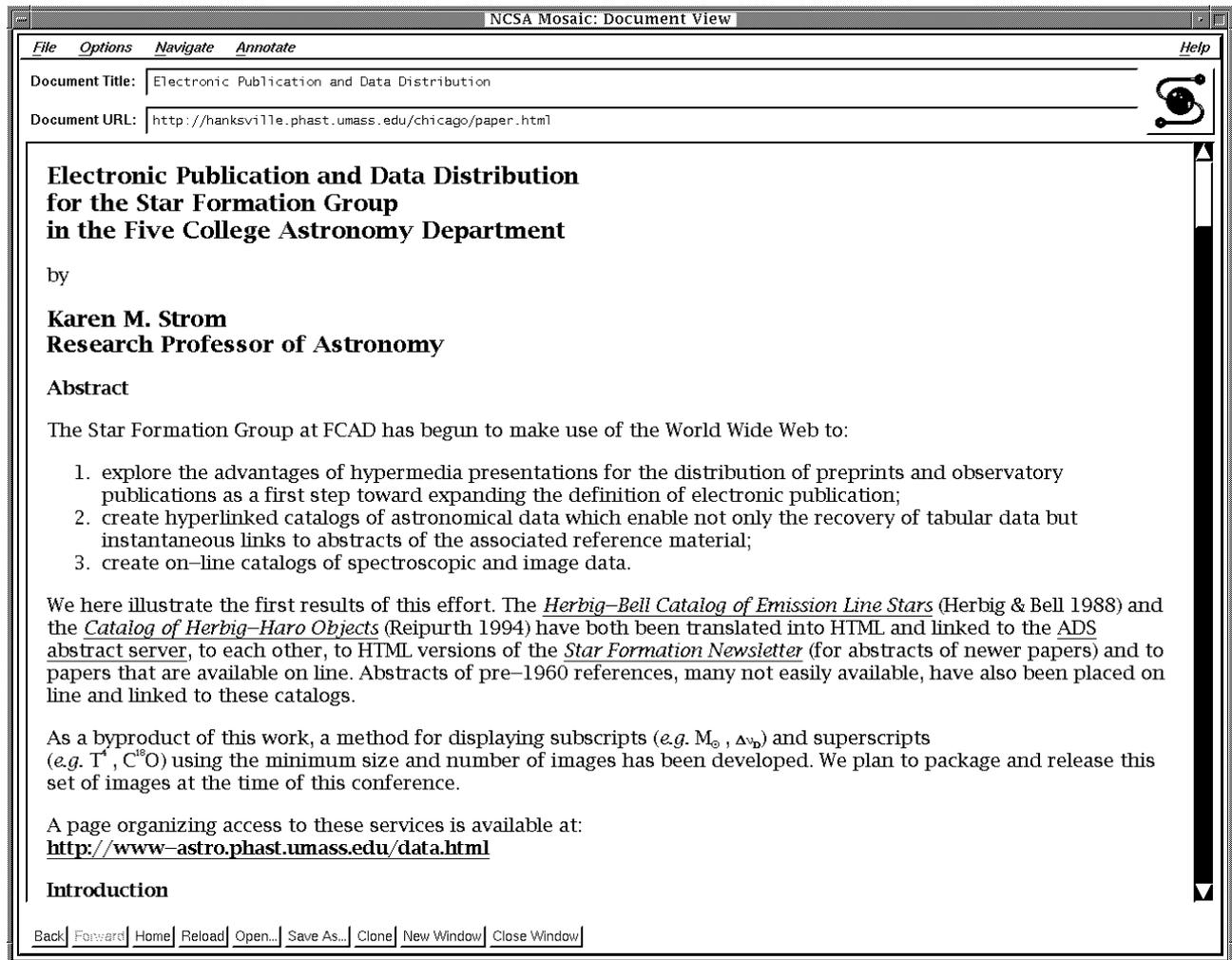
the text on the line, balanced on the subscript *e.g.* T_{eff} . Of course, because these expressions were inserted in the text as images, the font would rarely match that of the rest of the text.

In response to these practical and perceptual problems, I developed an image library of the most commonly used subscripts, superscripts and mathematical symbols used in the manuscripts which I had converted to HTML. This library was designed with several considerations in mind.

- The symbols should blend as inconspicuously as possible with the text.
- The total size of the images to be transferred should be as small as possible to minimize load time and network traffic.
- The reading of an entire paper should be accomplished without need to discard and reload either symbol or figure images if at all possible.

The first item on this list concerns the fact that, when the entire expression is captured as an image, with default font settings, as in \LaTeX2HTML , the expression tends to look out of place in the rest of the text generated by the browser, no matter which font is selected. However, if only the sub- and superscripts are generated as bitmaps, this problem is greatly alleviated. Many times letters are the main symbols used and are therefore simply part of the client generated text. The sub- and superscripts are always generated in another font and are much smaller so that few differences in the structure of the character are possible or noticeable.

The last two items take into consideration the default image cache size of most browsers, and the fact that the least recently accessed image will be the first discarded. The typical size of a 2 – 5 color image used as a figure in these preprints is 4 – 8 kilobytes. The typical size of a bitmap to be used as a sub- or superscript is 0.06 kilobytes. Thus our third consideration should easily be met.



These three considerations thus work in concert to provide a clean-looking, easily readable paper, taking minimal time to load each page, thus allowing browsing as well as reading in depth without forcing the scientific community to adopt another, temporary, shorthand for complex mathematical expressions or simply delaying the onset of electronic publication of journal papers.

The library includes the Greek letter set previously made available, the numerals, the entire alphabet, both capital and lower case, and some special symbols and letter combinations of common use in astronomy. To implement this set of transparent GIFs, *perl* scripts were written to preprocess the manuscripts to insert the in-line images as required. $\text{\LaTeX}2\text{HTML}$ is then used to convert the manuscript to HTML pages and a post-processing *perl* script is used to clean up the few things that may have been affected by $\text{\LaTeX}2\text{HTML}$. The manuscript is then ready for insertion of the images, tables and references.

As a sidelight, I note that the display of subscripts is a simple use of the HTML tag for displaying an in-line image. The unexpected *subscript* effect occurs because, in this case, the image is smaller than the text height. Thus

when you specify

```
<IMG ALIGN=MIDDLE ALT="_sun"
      SRC="/kicons/smsun.gif">
```

the middle of the small image is aligned with the base line of the text. Thus a subscript, M_{\odot} , is displayed. This method is also used to place the Greek letters having trailers properly on the baseline. The images are filled out with sufficient blank pixels below the letter that setting `ALIGN=MIDDLE` will properly locate the letter. I also urge users of this library to always make use of the `ALT = "xxxxx"` option so that people accessing your files with **Lynx** may more easily read your pages. This is easily accomplished when the substitutions are made in the manuscript by the *perl* script.

Although newer versions of $\text{\LaTeX}2\text{HTML}$ allow the option of conversion of Postscript figures, at a user-specified scale, into transparent GIFs in-lined into the text, I prefer to exercise more control over the scaling and appearance of the figures because the figures were initially designed for display on paper.¹ I display the file using *ghostview* and then scale the image to the desired size, then capturing the image using *xv*. For some images, in particular, gray-scale

¹ See my online tutorial on using images on the World Wide Web at:
<http://hanksville.phast.umass.edu/~kstrom/tutorials/images.html>

and full color Postscript files, ImageMagick may be better for this conversion process. I have chosen to place the scales, axis labels and other labeling text in a deep blue to separate it from the text. However, although the figure on the printed page may be very small, and we are not constrained by space considerations on the HTML page, the spatial resolution available to us is still less than that in the paper representation. Therefore I have chosen to make use of the fact that color is an option available at no extra cost to the server. When different line styles or point styles are called for, different colors are substituted by making use of *xpaint*. When the desired changes are completed, the file is converted into a transparent GIF and placed in the HTML document at the appropriate location. There is much more flexibility in figure placement in HTML documents as one is not forced to maneuver within fixed page sizes. For this reason, figures may be more reasonably located with respect to their textual references [2].

There are certainly some very complex figures for which this technique will not be appropriate. In that case a thumbnail or postage stamp image can be made, either by down-scaling the entire image or by cropping a recognizable section from the image for insertion into the text. This smaller image can then act as a link to the Postscript (or GIF) high resolution figure.

Tabular material presents different problems. Small tables can be placed within an HTML document by using the

<PRE> tag. Larger tables are only practical, at the moment, as Postscript files linked to the text at the appropriate point (but see the Catalogs section below). The use of HTML 3 should allow easy introduction of tabular material into the text. Very large tables can be associated with papers in other formats more appropriate to the data.

The Ph.D. theses can be found at:

<http://decoy.phast.umass.edu/>

The preprints from the star formation group may be found at:

<http://www-astro.phast.umass.edu/sfpreprints.html>

The library of transparent GIFs may be found at:

<ftp://hanksville.phast.umass.edu/pub/misc/newgifs.tar.Z>

A more complete paper describing their use is found at:

<http://hanksville.phast.umass.edu/chicago/paper.html>

References

- [1] Drakos, N. The manual for L^AT_EX2HTML is available from:
<http://cbl.leeds.ac.uk/nikos/tex2html/manual/manual.html>
- [2] Tufte, Edward R. 1983, *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT.