Astronomical Data Analysis Software and Systems XIV **P1.2.6**
ASP Conference Series, Vol. XXX, 2005
P. L. Shopbell, M. C. Britton, and R. Ebert, eds.

The CIAO Website Unveiled

Elizabeth C. Galle, D. J. Burke, Christopher Stawarz, Antonella Fruscione

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street Cambridge, MA 02138

Abstract. The Chandra Interactive Analysis of Observations (CIAO) website is the primary resource for users of the CIAO software package. Several hundred pages of content are written in XML, allowing for a baseline set of text from which many types of documentation may be created. Development and production (i.e. public) HTML versions of the site are generated from the XML via conversion scripts and XSL stylesheets. We present an overview of the back-end of the CIAO website, including custom markup tags, stylesheets, and CSS. The success of the project led to the use of this system in maintaining five websites at the Chandra X-ray Center (CXC).

1. Introduction

The CIAO software package[1] evolves relatively quickly, with an average of one major release and several smaller "patch" releases per year. The documentation needs to be updated at least as quickly as the software, and often on a much shorter timescale. By maintaining all of the documentation online, we are able to keep it up to date in real time and make modifications available to the user community immediately.

The result of the CIAO documentation effort is a set of websites that contain several hundred pages of current content, as well as archived information for previous software versions.

Maintaining this information led us to create a customized content management system which would allow multiple authors to write documents, review them in a closed development environment, and then post them on the live site. Here we present an overview of the back-end of the publishing system and discuss some of the benefits from its use.

2. Architecture

The publishing system (hereafter referred to as "CIAODOC") used to create the CIAO website consists of several interrelated components which control content,

[1] `http://cxc.harvard.edu/ciao/`

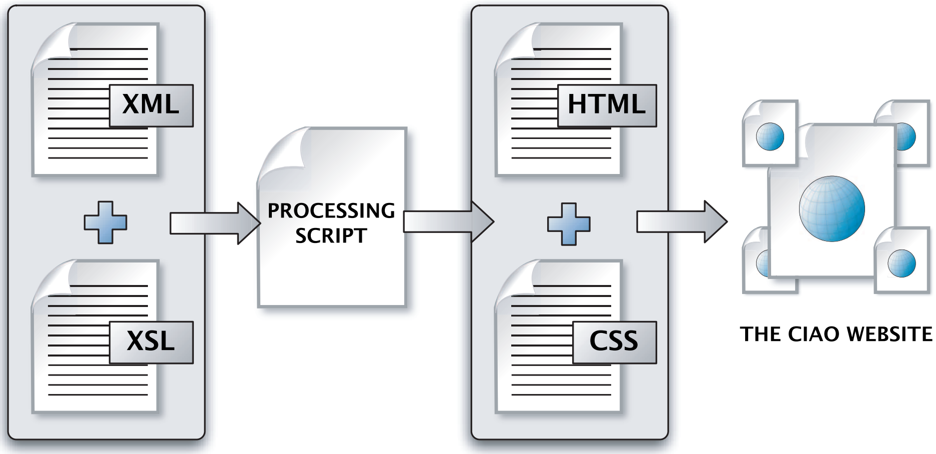

Figure 1. The CIAODOC processing system.

layout, and style. The following process is currently used to maintain five different sites (including one internal) at the CXC.

As shown in Figure 1, CIAODOC takes the XML source code, transforms it with XSL stylesheets (via a processing script), and produces HTML pages. CSS is used to apply color and other styles, completing the CIAO website.

2.1. XML Source Code

All of the CIAO webpages are written in XML (Extensible Markup Language)[2], a versatile text format from which many types of documentation may be created. The customized XML syntax is similar to (X)HTML and has many tags in common with it: paragraphs (`<p>`), `<table>`s, `<div>`s. While the basic structural elements are the same, a number of new tags were defined to simplify the writing, notably for the creation of links.

CIAODOC uses over twenty different custom link tags in place of the HTML anchor (`<a>`) tag. The convenience of these link tags is that the author is not required to give a complete path, just the name of the target page. CIAODOC determines the path to that page from the current location and creates the link accordingly. A few examples of link tags used in the XML source and the corresponding HTML destination are shown in Table 1.

Another customization is the creation of different page types. Each XML file begins with a tag that specifies the page type, such as `thread` or `dictionary`. The tag is required in order for CIAODOC to apply the correct transformations during the processing (see §2.2.).

The files are stored in a directory structure separate from the web pages, so that they are not visible to the user. SCCS (Source Code Control System) is used for version control, a vital feature in a multiple author environment. SCCS prevents more than one person from editing an XML file at a time, and

[2] `http://www.w3.org/XML/`

Table 1. Sample Link Tags.

XML tag	HTML link
`<aguide>`	http://cxc.harvard.edu/ciao/guides/
`<ahelppage>`	http://cxc.harvard.edu/ciao/ahelp/
`<threadpage>`	http://cxc.harvard.edu/ciao/threads/
`<why>`	http://cxc.harvard.edu/ciao/why/

maintains a history file for each XML file. The history file tracks the edits to the XML and makes it possible to retrieve any previous version of the document.

2.2. XSL Transformations

XSLT (Extensible Stylesheet Language Transformations)[3] is a language used to transform XML into an output format, whether it be HTML, text, or new XML. The CIAODOC XSLT consists of three dozen stylesheets used to process the XML into HTML. There are specific stylesheets to handle different page types, as well as many all-purpose stylesheets that are used in processing every type of page.

For example, when processing an analysis thread[4], CIAODOC uses one stylesheet which creates thread-specific structure (overview, table of contents, section blocks, summary), while another stylesheet is employed to process all the links in the document. Depending on how complex the content is, additional stylesheets could be called to perform other transformations.

Some error handling is also built directly into the XSLT to prevent authors from omitting required information, such as link content.

2.3. Processing Script

A Perl script which calls xsltproc[5] is used to process the XML files. The script serves two main functions:

1. **HTML creation:** the XML source code is converted to HTML by applying XSLT stylesheets (see §2.2.)
2. **File management:** the newly generated files are copied to the designated output directory.

This process is known as "publishing" the pages. By default, the script publishes the page to a development server so that it may be reviewed for accuracy. Once it is ready to be made public, a command-line switch is used to publish it to the live server.

---

[3] `http://www.w3.org/TR/xslt`

[4] `http://cxc.harvard.edu/ciao/threads/all.html`

[5] `http://xmlsoft.org/XSLT/xsltproc2.html`

The script may also be run on non-XML files (such as images, text files, or postscript files); in this case the files are simply copied to the correct directory.

2.4. CSS

Once CIAODOC has created the HTML version of the pages, Cascading Style Sheets (CSS)[6] are applied. CSS is primarily used to control the look of the navigation bar (e.g. the signature blue background), as well as the behavior of links when they are selected. It also adds highlighting in other areas of the site, and handles some minor positional information. The CSS file for the CIAO website can be viewed online:

`http://cxc.harvard.edu/ciao/ciao.css`

CSS is not used more extensively in part because Netscape 4.7x - a browser still favored by a number of our users - has substandard support for CSS.

2.5. Creating Hardcopies

Since the CIAO documentation is exclusively released online, we provide two convenient methods of obtaining hardcopies of the webpages.

- **PDF files:** after the HTML pages are generated, CIAODOC creates PDF files of them in Letter and A4 formats. Links to the PDF files are provided at the top and bottom left of every webpage for download.
- **Print-specific CSS:** just as stylesheets are used to format the pages for display in your browser, we use a stylesheet that relays print formatting instructions to modern browsers. This stylesheet is applied when you send a page to the printer or use the "print to file" option.

In both cases, some minor modifications are made to the page content before creating the hardcopy. The URL of the page is included at the top of the document, and details that aren't useful in print, such as the navigation bar, are omitted.

3. Validation

The XML source code is checked before publication to make sure that it is both well-formed and valid. A "well-formed" document is structurally correct (e.g. all tags are closed and properly nested), while a "valid" one contains all the required elements for that particular document type. Both of these criteria must be met in order for CIAODOC to handle the code properly. A RelaxNG[7] schema which defines the structure of the document, similar to a document type definition (DTD) for an HTML page, is used to examine the XML file prior to processing.

[6] `http://www.w3.org/Style/CSS/`

[7] `http://www.relaxng.org/`

CIAODOC is designed to create HTML that meets the "4.01 Transitional (no frames)" DTD. The live HTML pages are periodically validated via the W3C Markup Validation Service[8] to ensure we are achieving this goal.

4. Conclusion

Originally developed with only the CIAO website in mind, the CIAODOC publishing system is now used to maintain four public CXC websites: CIAO, *Sherpa*[9], ChaRT[10], and CalDB[11].

Since all of these sites are built by CIAODOC, it is easy to apply the same layout and color schemes (i.e. "branding") across the board. After a user is introduced to one of the sites, he will be able to recognize instantly that he is browsing one of the CXC software webpages on subsequent visits.

Along with the consistent user experience afforded by this system, there are a number of benefits to both the developers and the users:

- Any page structure beyond basic organization is determined by the publishing scripts rather than being hardcoded into the site. This allows us to make site-wide changes easily, instead of editing every individual XML file.
- Another benefit of separating the content and presentation is the ability to maintain a consistent look and feel in a multiple-author environment. An author can control the basic organization of the page content, but the layout is determined by the XSL stylesheets.
- Forward compatibility: producing valid code helps ensure that the website will continue to work correctly as newer browsers are developed.

Acknowledgments. This work was supported by the Chandra X-ray Center under NASA grant NAS8-03060.

[8] `http://validator.w3.org/`

[9] `http://cxc.harvard.edu/sherpa/`

[10] `http://cxc.harvard.edu/chart/`

[11] `http://cxc.harvard.edu/caldb/`