

+ Page 5 +

-----  
Price-Wilkin, John. "Using the World-Wide Web to Deliver Complex Electronic Documents: Implications for Libraries." The Public-Access Computer Systems Review 5, no. 3 (1994): 5-21. To retrieve this file, send the following e-mail message to listserv@uhupvml.uh.edu: GET PRICEWIL PRV5N3 F=MAIL. (The file is also available from the University of Houston Libraries' Gopher server: info.lib.uh.edu, port 70.)  
-----

## 1.0 Introduction

The World-Wide Web (also called the Web) is a very promising tool for libraries to use to explore the delivery of rich and complex documents. [1] Nevertheless, there are many limitations in the Web's HTML markup language and the ability of Web servers to deliver structured information. This paper explores the benefits and limitations of the Web in the context of several projects taking place at the University of Virginia, both in the Library and in the University's Institute for Advanced Technology in the Humanities. A gateway between the Web and the SGML-based PAT system that helps to overcome the Web's inherent limitations is also described.

## 2.0 SGML and TEI

The most worthwhile products that libraries can buy are ones that conform to standards and are not tied to a specific software package or operating system. These are the only products with enduring value. Certainly, there are exciting electronic resources being produced for specific software packages and operating systems, but the extent to which libraries can build collections of hypertext resources that are usable in the future will depend entirely on the conformance of their resources to true national and international standards. The most important standard for this discussion is SGML, a standard designed to express the organization of documents and to accommodate even the most complex multimedia materials.

+ Page 6 +

A brief (and admittedly superficial) discussion of SGML and the Text Encoding Initiative may be helpful. SGML (Standard Generalized Markup Language, ISO 8879) is a standard approved by the ISO for the descriptive markup of documents. The language of SGML is sufficiently flexible that the sense of "document" has been expanded to include coordinated time-based elements of hypermedia (e.g., animated dance, music, and character-based score and choreography moving in synchrony at a pace controllable by the user). SGML is not a tag set: there are no pre-set tags. Instead, SGML is a set of rules (or a grammar) for articulating that vocabulary. These rules are sufficiently rigorous so that specialized software can check the validity or conformance of a document. The specification of that grammar is a DTD (Document Type Definition); the DTD can also function to document many decisions about the organization of a text. Without that validity--i.e., without being parsed against a DTD--the document is not SGML encoded, although it may share many of the

characteristics of SGML.

For our work at Virginia, the most notable of these characteristics has been the descriptive nature of the tagging. Rather than saying that an element of the text appears in bold, 17 point Helvetica, centered at the top of a new page, we use the tags to define the function of a textual element (e.g., a title). The tag set used must necessarily elaborate the elements of the texts we see in an academic environment: a tag set designed for articles or documentation, for example, will omit important elements needed for encoding poetry. To serve those needs, the Text Encoding Initiative (or TEI) has published a set of guidelines for the application of SGML to texts in the humanities. Functions of the text or hypertext, expressed descriptively and with a standard language, are freed from the constraints of a specific software package or application. SGML-encoded works can serve a variety of functions, depending on the user's needs and available software.

+ Page 7 +

### 3.0 The Potential of the Web

The Web uses a client/server architecture. Sophisticated Web clients, such as Mosaic, offer an exciting sense of the possibilities of electronic publishing on the network. Several revolutionary concepts that have been awaited with anticipation are incipient in all aspects of that relationship between client, server, and publication. These characteristics are:

- o Open systems--the ability to make resources available to a variety of operating systems and a variety of applications is evident throughout the Web. Computers running X Windows, Microsoft Windows, and the Macintosh System 7 all participate equally. In addition to Mosaic, other clients, such as Cello and OmniWeb, are available. Multimedia tools, such as image viewers, are a matter of personal choice.
- o Standards--given the Web's use of HTML, the importance of standards is heightened, and HTML is inexorably moving toward greater expressiveness and greater conformance to the SGML standard.
- o Distributed information--the notion of a universe of distributed information, scattered throughout the Internet while being conceptually linked to other information, is becoming a reality through the use of the Web.

### 4.0 Representative Web Projects

Over the past two years at the University of Virginia, faculty and staff involved in several projects began to develop a variety of electronic materials using the SGML standard. Partly this was to serve already apparent needs, but it was also to take advantage of the potentials of electronic publishing. While the Library's Electronic Text Center and, later, its Digital Image Center began to develop skills in creating electronic materials in standard formats for networked access, scholars at the Institute for Advanced Technology in the Humanities undertook the daunting task of composing advanced, standards-based electronic

research materials without having the tools with which to publish these materials. With the introduction of Mosaic, the Web was quickly seen as a way to deliver these materials, and, with relative ease, large bodies of SGML-encoded material were converted to HTML for Web access. In order to focus on particular aspects of those projects, the following example projects are divided into sections on editions, history, image archives, and instruction.

+ Page 8 +

#### 4.1 Editions

In general, the Web offers creators of editions of literary or other works the ability to represent a vast, interconnected web of scholarly resources in a variety of different ways. The user might view the resources simply, as in an edition of a work without the introduction of a critical apparatus. A more complex approach is also possible, with the user following the critical apparatus at every turn. And finally, a rich and scholarly approach is possible, allowing the user to view manuscript (or printing) evidence or to examine the editor's assessment of the evidence by comparing high-quality scans of original pages to the marked-up transcriptions. With proper markup, an edition can be viewed in as many ways as the reader desires. It can be a variorum, a study edition, a critical edition, or historical evidence. The form the edition takes is defined by the user's needs or preferences.

##### 4.1.1 British Poetry

The British Poetry Archive documents are perhaps the simplest of those discussed here. (The project's URL is <http://www.lib.virginia.edu/etext/britpo/britpo.html>.)

The two texts now available were transcribed by students in Jerome McGann's graduate courses. In addition to the SGML-encoded text itself, each work includes material such as introductions, notes, and glosses as well as high-quality digital facsimiles of pages from the original editions. The materials are freely available on the Internet, and Mr. McGann hopes that others will contribute to the archive. These texts represent the simplest of the hypertext editions available on the University of Virginia's Web, with supporting materials providing potential deviations from an otherwise linear progression. The texts were encoded in TEI-conformant SGML with the assistance of the Library's Electronic Text Center, and they were then converted to HTML for the purpose of making them available on the Web.

+ Page 9 +

##### 4.1.2 Dante Gabriel Rossetti

To date, the most fully developed project is Jerome McGann's ongoing edition--or archive--of the works of Dante Gabriel Rossetti. (The project's URL is <http://jefferson.village.virginia.edu/rossetti/rossetti.html>.) According to McGann, the Rossetti archive is:

a hypermedia environment for studying the works of the Pre-Raphaelite poet and painter D. G. Rossetti (1828-1882). The archive is a structured database holding digitized

images of Rossetti's works in their original documentary forms. Rossetti's poetical manuscripts, early printed texts --including proofs and first editions--as well as his drawings and paintings are stored in the archive, in full color as needed. The materials are marked up for electronic search and analysis, and they are supplied with full scholarly annotations and notes. [2]

The organization of the archive is designed to capitalize on the uniquely intertwined nature of Rossetti's artistic process, linking image to text and text to image. When Rossetti accompanied a painting by sonnets, the poems are included in the archive along with an image of the painting. When Rossetti illustrated a poem with a painting, an image of the painting is included. Since Rossetti frequently designed his own editions, electronic versions of his print works, with linked text and images, are also available. McGann describes the difficulty of studying Rossetti's works in a traditional print environment, and then sets about trying to overcome those difficulties by melding the resources in a way that allows the reader to follow the threads of art, poetry, or translations without losing access to the other materials.

#### 4.1.3 Piers Plowman

The third project was begun in the 1994-95 academic year by one of the most recent Institute fellows, Hoyt Duggan. (The project's URL is <http://jefferson.village.virginia.edu/piers/archive.goals.html>.)

+ Page 10 +

Mr. Duggan, an accomplished editor of Middle English texts, created an edition of the Piers Plowman B text using the Web. More in the model of the traditional scholarly edition, Mr. Duggan's project brings together transcription and facsimile to resolve vexing editorial problems. When the scribe uses an abbreviation to represent a letter combination (e.g., a barred "p" for "pre"), the reader typically wants the editor's best judgement in rendering what was intended (i.e., "pre"). Many of those decisions deal with unambiguous evidence, and some with less certain evidence. Through SGML, both the suspension or abbreviation is registered as well as the reading of the character.

To the greatest extent possible, digital facsimiles of all seventeen surviving manuscripts will be included. With facsimile evidence, it is always possible to return to something resembling the original document to evaluate the editor's decision. Duggan has also found that it is possible to create extremely high-resolution images that, with enlargement and other digital treatments, can reveal important new information about the original composition.

#### 4.2 History

With new technological tools, historians are offered both challenges and opportunities. Electronic resources allow them to blend evidence and interpretation in ways that help both student and researcher. A simple approach in using the materials is possible, where the reader follows the argument without examining evidence. It is also possible for the reader to examine the

methodology of the researcher, either to scrutinize the research or to be instructed in the methodology of research. The process of bringing evidence and interpretation together brings challenges of immense proportions. For example, the role geography plays in defining an event can be brought to bear on the problem, but it may involve the use of sophisticated systems of geographic analysis. Two projects at the Institute have used many diverse resources to explore their topics, incorporating nineteenth-century Census data, geographic models, and animated sequences.

#### 4.2.1 Ayers (Valley of the Shadow)

Edward Ayers, a historian of the Civil War and the Reconstruction, was one of the Institute's first two fellows. (The project's URL is <http://jefferson.village.virginia.edu/vshadow/vshadow.html>.) According to Ayers, the project:

+ Page 11 +

interweaves the histories of places on both sides of the Mason-Dixon line. It is the story of two communities relatively close to one another, sharing considerable prewar characteristics and similar experiences in the war itself. There was one area in the United States for which that was most clearly the case: the Great Valley that stretched from Pennsylvania, through Maryland and Virginia, into Tennessee. [3]

Ayers focuses on two towns--Staunton, Virginia and Chambersburg, Pennsylvania--as representative communities from that Valley that served as such an important economic, cultural, and military locus of the War. The Web serves the historical ends by balancing narrative--a filtering or interpretation of evidence--with the presentation of that evidence. Ayers has described one dilemma of the historian as a tight-rope act between providing access to evidence and creating an organizing argument that does not also obscure that evidence. His approach, providing the deepening layers of evidence as "rhizomes" beneath the surface of narrative, has been well-supported by the Web.

#### 4.2.2 Dobbins (The Forum at Pompeii)

Dobbins, a classical archaeologist, reconstructs Pompeii from archaeological evidence in a virtual space to advance his argument. (The project's URL is <http://jefferson.village.virginia.edu/pompeii/page-1.html>.)

He uses computer-aided design (CAD) tools to bring precision to his reconstruction. Animation is being added to the CAD representations to provide a three-dimensional perspective of buildings and space. Structures that are normally seen in isolation from each other are assembled in a total vision of Pompeii that may suggest a degree of planning and coordination.

#### 4.3 Image Archives

The Digital Image Center's image collections can be seen as passive collections of standards-based images. (The project's URL is <http://www.lib.virginia.edu/dic/class/arh102>.)

The image collections are organized to reflect the focus of an individual class or an art exhibit. All of the images are

TIFF files subjected to JPEG compression. As such, they can be examined with a variety of image tools, ranging from simple viewers to software with analytical capabilities. Most importantly, the tool used is largely the choice of the user. As a result of planning and philosophy, all images are durable enough to stand close scrutiny: they were scanned in 24-bit color at a sufficiently high resolution to be enlarged several times without significant degradation.

+ Page 12 +

The most developed collection is representative of this archival philosophy. William Westphal's graduate architectural history course on urban form includes hundreds of architectural images, primarily from the Italian Renaissance, organized around his lectures. Students can access these resources at all times over the network as well as in a closed classroom environment designed to efficiently access the images. Since they were scanned at high resolutions, the images compare favorably with the original slides, and they can be examined closely on screen. The original slides have frequently degraded or had imperfections that were corrected in the scanning process.

#### 4.4 Instruction

The final project demonstrates the instructional capabilities of the Web. (The project's URL is <http://www.lib.virginia.edu/etext/scanner.html>.)

Using the Web to provide access to training materials has many strengths. It gives variation to what would otherwise be a flat, linear document. The document is dynamic and can easily accommodate other elements as they are created by staff.

Scanning text is one of the most repetitive training operations provided in the Electronic Text Center. Unlike searching electronic texts, where every research need may entail a different approach and different training needs, many of the scanning decisions are generalizable and can be represented in a training document. The project's instructional Web pages on scanning were designed to reduce the amount of staff intervention and give a greater degree of freedom to users.

#### 4.5 Evaluation of the Projects

While the majority of the projects discussed here could be supported by numerous stand-alone, operating-system specific hypertext products, the Web has several advantages.

The projects' electronic resources are widely available on the Internet, and users can access them on a variety of computer platforms, regardless of the fact that the Web server is running on a UNIX computer. (Attractive graphical Web clients, such as Mosaic and OmniWeb, are available for Macintoshes, IBM-compatible computers using Microsoft Windows, UNIX computers with X Windows, and NeXTs.)

+ Page 13 +

Another key advantage is that the source material for the editions either conforms to or is in the process of being composed using international standards; it is marked up to suggest the functional characteristics of the collections, rather than their representational characteristics. Elements, such as

titles, quotations, and headings, are marked to suggest their functional role in the document, rather than any presumed display value. Displays depend instead on the capabilities of the user's software, which utilizes the functional characteristics of the elements to determine how to present the information.

This reliance on functional--not representational--characteristics means that the same materials can be used in a variety of different ways, supporting the creation of editions with other software packages (e.g., Electronic Book Technology's DynaText), use with different analytical tools (e.g., morphological parsers), and access through different database schemes (e.g., text-specific systems or relational database managers designed for images). A high degree of flexibility, viability, and multi-platform access can be maintained.

Each of the mentioned editions and historical analyses was first composed in a very rich SGML format that was designed to discriminate between the functional characteristics of low-level elements. They were subsequently converted (as automatically as possible) to static HTML versions for use with the Web. Elements, such as discrete descriptive bibliographic characteristics, become simple list items, and most complex prose and verse elements are reduced to paragraphs and line breaks. After this conversion, it was discouraging to see that richness disappear, but the original document remained unchanged.

There is a continued expectation by the scholars who created these resources that better tools will be developed to tap the inherent complexity of these materials. The standards-based format of the materials ensures that these scholars will be able to take advantage of these new tools when they become available.

## 5.0 The Web as an Authoring and Document Delivery Environment

The authoring and document delivery capabilities of the Web are significantly limited for documents of even moderate complexity. Authoring for the Web is usually done in HTML. HTML has many virtues, not least of which is its striving for expressiveness and SGML validity. It is, however, an impoverished tag set with little ability to reflect the complexities of most of the documents discussed earlier, despite their being offered through the Web. It is important to note that the Web is a limited document delivery environment. Its inability to recognize or use structural features of documents forces unpleasant administrative decisions that will likely restrict the later use of these documents.

+ Page 14 +

### 5.1 HTML's Lack of Expressiveness

The range of HTML tags available to users is limited. In contrast to the hundreds of tags made available by the TEI guidelines, roughly two dozen tags are made available in HTML. While HTML will be expanded with HTML+ to give greater precision in areas such as tabular data, HTML+ cannot be expected to provide the breadth needed to support literary and historical documents, or even to support standard journal literature.

This lack of expressiveness and insufficient breadth of tags also leads to the author's inability to differentiate important elements with HTML. In HTML, the same small set of tags is necessarily used for diverse sets of elements. For example, the <BR> code (line break) is used for verse lines, table elements,

stanza divisions, dramatis personae, and many features. Authors are also left with little ability to represent the structural organization of a document. Where the author wishes to define a bounded segment of text, such as a stanza or chapter, no tag is available for this purpose. Instead, authors rely extensively on dividing documents into files representing major structural divisions. Elements that are normally defined as structural tags in SGML, such as the paragraph (or <P>) tag, are not defined by HTML in a way that reliably defines the contents of a paragraph. This paucity of tags in HTML results in the author of any document of moderate complexity using many tags to effect a desired appearance, rather than to characterize the content. This type of tagging confuses function and appearance.

The inability of HTML to represent complexity is often closely linked to the inability of Web servers to provide access to complex representations of documents. This inability is fundamentally linked to the notion of structure. Where structural distinctions exist in the markup language, there is no inherent ability in the Web to deliver that individual element. So, for example, HTML defines glossaries and glossary entries, but, in order to provide access to an individual glossary entry from a hypertext link, the server must send the entire file (i.e., the file containing the glossary) to the user. Smaller glossaries cause few problems, but this makes providing access to individual "glossary" entries in a document such as the Oxford English Dictionary, where all 500 MB would be transferred across the network, effectively impossible. While Web browsers are intelligent enough to move automatically within the file to the chosen glossary entry, the file transfer paradigm is impractical for large-scale information delivery. Given this, it must also be pointed out that there are very few HTML tags that define structural relationships. Structures such as chapters, sections, or poems are not represented.

+ Page 15 +

The Web's deficiency with regard to structural features leads to decisions with serious negative administrative consequences. Because the Web does not include structure awareness in its protocol and because HTML markup provides so little support for structural representation of features, the author and the administrator are forced to fragment documents into a sets of reasonably sized components. In converting the ARL book University Libraries and Scholarly Communication (URL: <http://www.lib.virginia.edu/mellon/mellon.html>) to HTML, I found that, using the Web and HTML alone, it was necessary to divide the dozen chapters into separate files. While this may not sound onerous, extending this practice to a large collection of documents--or even a small collection of large documents--would be very difficult. An HTML version of the OED would become a set of 300,000 files. Chadwyck-Healey's English Poetry Database would become either 2,500 files (if the administrator wished to provide access at the volume level) or 65,000 files (if access to individual poems were supported). Even this severe approach does not solve needs that might arise for substructures, such as quotations and definitions within the OED or specific stanzas within a poem.

## 5.2 Overall Limitations of HTML

For documents of limited complexity, HTML is an effective



authoring environment; however, it seriously limits the ways in which a more complex document or a set of documents can be used. No differentiation of important elements (e.g., stanzas and subdivisions of prose) can take place, and it will be necessary to upgrade the coding of HTML documents within the year.

The Web also lacks inherent document management or document access capabilities. In part because of the limitations of the markup language and in part because of the design of the protocol, there is a paucity of structure represented and no structure recognized. I emphasize "inherent," however, because the Web also provides a gateway capability that can more than compensate for this deficiency.

## 6.0 Exploring Alternatives

I have been developing a gateway from the Web to an indexed collection of texts in an SGML-aware system to take advantage of the complexity of the documents and yet make them available through the Web. The texts are nearly all in fully validated SGML tag sets, each with significant expressiveness. In contrast to an HTML collection, potentially consisting of many files representing the many component parts of the collection, each text is a single file with as many as hundreds of thousands of structural components.

+ Page 16 +

## 6.1 Collections

Three diverse examples are provided to help understand the nature of the collections used in the gateway.

### 6.1.1 University of Virginia Middle English Collection

The Middle English collection assembled by the University of Virginia's Electronic Text Center is approximately thirty texts in a single file. (The collection's URL is <http://etext.virginia.edu/Mideng.query.html>.)

Texts vary in size from several dozen pages to several hundred pages. One of the Library's smaller collections is approximately 11 MB of raw text, but it grows as new materials become available. The markup language used is SGML complying with the Oxford Text Archive's DTD, a tag set that will eventually represent a valid subset of the TEI DTD. The tags differentiate major structural elements, such as tales in the Canterbury Tales, bibliographic elements, and elements of composition (e.g., verse lines, stanzas, and paragraphs). Markup is rich enough to support a wide range of analytical requirements, and the texts have been made available for the purpose of analysis to the University of Virginia community for much of the past two years. With the permission of Open Text, the Oxford Text Archive, and creators of individual texts, access to this collection is unrestricted. It can be accessed in a variety of ways, including the Web.

### 6.1.2 Chadwyck-Healey English Poetry Database

The Chadwyck-Healey English Poetry Database is purchased on tape from the publisher and made available indexed by PAT. Access to this collection is restricted to a consortium of five universities in Virginia. As yet incomplete, the collection

currently consists of nearly 1,600 works with more than 64,000 poems and 233,000 pages. The raw text is relatively large (340 MB), but, indexed with PAT, searches usually yield results in less than one second. The SGML used with the English Poetry Database is a very rich set of tags designed in consultation with a TEI representative. It is more than adequately expressive about the poems, including structural markup for poems, poem divisions such as stanzas, lineation, and attributes such as whether rhyme is used.

+ Page 17 +

### 6.1.3 Oxford English Dictionary

The Oxford English Dictionary is the largest and arguably the most complex resource made available through this service. The 570 MB document contains approximately 300,000 entries, many with more than fifty subelements. Strictly speaking, it is not in SGML form because it has not been validated against a DTD. The electronic version was, however, designed to take advantage of SGML's characteristics, and it significantly benefits from the file's structural and descriptive markup.

## 6.2 Web to PAT Gateway

I have constructed a gateway between the Web and the more sophisticated SGML texts using the Web's CGI (Common Gateway Interface) and PAT, an SGML-aware text retrieval program. Text is returned from PAT to the Web in the richer SGML, and it is converted on the fly to HTML, primarily using HTML to control the appearance of the text on the screen. This gateway is being documented elsewhere (URL: <http://sansfoy.lib.virginia.edu/pub/www-to-pat/>), but several facets are relevant to this discussion.

### 6.2.1 Expressive Representation of Text is Retained

The original unmodified texts are accessed through the gateway without compromising the expressiveness of the original markup. Although the sophisticated SGML markup is dynamically rendered as HTML as the user retrieves results, the text remains in the original rich SGML form behind the Web representation. Decisions about the way that the fuller tag set maps to HTML are registered in filters, and, as HTML becomes more expressive, a better match between the original tags and the HTML can be made.

### 6.2.2 Simple Queries and Simple Access

Users need not be familiar with PAT's query language to search texts and take advantage of the structural characteristics of the more expressive markup. A word or phrase search returns keywords-in-context (KWIC) views to the user, from which a view of larger context is possible. Eventually, this process may lead the user to retrieval of entire sections (e.g., chapters or acts). All expanded views are made from hypertext links that initiate structural retrievals such as "the chapter that includes this search result."

+ Page 18 +

### 6.2.3 Menu-Driven Structural Queries

It is possible to facilitate complex queries through menus. For example, in the OED, the word lookup function facilitated by the Web includes queries such as: "give me entries that include my word within the Lookup field of the Headword Group field," or "give me entries that include my word in the Variant Form field." The user is not aware of the complexity of the query taking place, but can modify the type of query by selecting different variations on the search menus. Boolean queries that ask for the intersection of document structures have been challenging to users employing command-line and analytically oriented interfaces. However, through simple fill-out forms and menu selections, queries such as "(stanzas including [word/phrase]) INTERSECT (stanzas including [word/phrase])" are executed without the user needing to understand the system's command syntax. While we also offer access through several complex, analytical interfaces (PatMotif and PowerSearch from Open Text as well as a locally developed VT 100 interface), most users can avoid these more complicated interfaces.

#### 6.2.4 Access to Structure

Finally, the administrator of a collection need not resort to fragmenting files to make it possible to provide access to the component parts of a collection. As mentioned earlier, an HTML approach to the OED would require us to divide it into 300,000 files. I was recently able to represent the dozens of parts, chapters, sections, and subsections of a voluminous SGML technical document through this strategy, making hypertext links and each component accessible by utilizing the fairly rich markup; however, the document remained a single file. Resource management is made more reasonable through a system cognizant of a file's structure.

+ Page 19 +

#### 6.2.5 Future Approaches

This strategy has many possibilities. Journal literature coded in SGML may be successfully accessed through this sort of strategy. For example, a journal run marked up according to the more elaborate Association of American Publishers DTD could return articles to the user through PAT queries. Another approach would facilitate browsing by recognizing the structural relationship of author and abstract to article, article to issue, issue to volume, and volume to collection. Throughout, the collection would exist as a single file, searchable across all articles by a single query. The collection would not need to be compromised by converting the articles to HTML, but would instead continue to remain in the more expressive AAP SGML format, filtered for display in the process of retrieving information. Through this strategy, the Web can be an effective means of accessing the original files in a fuller SGML, without resorting to fragmenting the material into files corresponding to the individual articles or even parts of articles. Similar strategies for books and documentation are possible.

### 7.0 What Does the Web Offer Libraries?

The Web is a complex system with great potential and serious limitations. We should use caution as we consider composing in

HTML: it is a short-term coding strategy. Documents composed in HTML will have limited expressiveness, and, because HTML is not yet stable, they are likely to need continuing enhancement to be used in the Web. There is much to be excited about with the Web: it is a viable system that suggests what electronic publishing on the Internet can be. We have lacked credible, demonstrable examples of standards-based, networked hypertext in the past, and the Web has changed that. There is a great deal of untapped potential in the Web. By exploiting the Web's ability to talk to other more sophisticated programs, we can begin to take advantage of that potential and make tomorrow's promise real today.

+ Page 20 +

A subtext of this article has been the importance of standards--both employing them in creating hypertexts and extending the Web to take greater advantage of them. Standards have been attractive to libraries because they help ensure long-term viability. However, as Jefferson remarked in 1790, standards are also an important key to information being generally useful, regardless of context:

Measures, weights and coins, thus referred to standards unchangeable in their nature . . . will themselves be unchangeable. These standards, too, are such as to be accessible to all persons, in all times and places. The measures and weights derived from them . . . are within the calculation of every one who possesses the first elements of arithmetic, and of easy comparison, both for foreigners and citizens, with the measures, weights, and coins of other countries. [4]

#### Notes

1. A version of this article was presented as a paper at the Yale Hypertext Conference, May 1994. An HTML version of the original speech, with active links to the resources discussed, is available via the World-Wide Web; URL: <http://sansfoy.lib.virginia.edu/pub/yale.html>.

2. Jerome McGann, *The Complete Writings and Pictures of Dante Gabriel Rossetti: A Hypermedia Research Archive* (Charlottesville, VA: Institute for Advanced Technology in the Humanities, University of Virginia, 1994). (Electronic document available via the World-Wide Web; URL: <http://jefferson.village.virginia.edu/rossetti/rossetti.html>.)

3. Edward Ayers, *The Valley of the Shadow: Living the Civil War in Pennsylvania and Virginia* (Charlottesville, VA: Institute for Advanced Technology in the Humanities, University of Virginia, 1994). (Electronic document available via the World-Wide Web; URL: <http://jefferson.village.virginia.edu/vshadow/vshadow.html>.)

4. Thomas Jefferson, "Public Papers," in *Writings* (New York: Literary Classics of the U.S., 1984), 410.

#### About the Author

John Price-Wilkin, Systems Librarian for Information Services, Alderman Library, University of Virginia, Charlottesville, VA

22903. Internet: jpw@virginia.edu.

+ Page 21 +

-----  
The Public-Access Computer Systems Review is an electronic journal that is distributed on the Internet and on other computer networks. There is no subscription fee.

To subscribe, send an e-mail message to listserv@uhupvml.uh.edu that says: SUBSCRIBE PACS-P First Name Last Name.

This article is Copyright (C) 1994 by John Price-Wilkin. All Rights Reserved.

The Public-Access Computer Systems Review is Copyright (C) 1994 by the University Libraries, University of Houston. All Rights Reserved.

Copying is permitted for noncommercial use by academic computer centers, computer conferences, individual scholars, and libraries. Libraries are authorized to add the journal to their collection, in electronic or printed form, at no charge. This message must appear on all copied material. All commercial use requires permission.  
-----