

The eight benefits described in this chapter were identified by the Tested teams through the process of developing their prototypes and business cases. CTG administered surveys and conducted personal interviews with all the participants to refine these findings into those categories that exhibited the highest level of consensus and impact. The quotations used within the descriptions of the benefits are taken directly from the Testbed interviews.

Information consistency

Information consistency refers to text, images, and other content remaining the same regardless of how and where they are presented. In other words, though the presentation may vary with the media, the content remains the same. This holds true whether it's a printed publication, a Web page, mobile device, or a word processing format—to name just a few.

Ensuring this consistency frequently involves managing several different formats and multiple source documents. A change at any one point requires changes at all other points. As the number of presentation formats and source documents increase, so does the percentage of errors and inconsistencies, since the original author does not always perform the changes. Ownership of the content and responsibility for maintaining consistency can become muddled in this process. As one Webmaster explained, ***“They [content developers] rely on us [technical team] because we have always done this, if the text doesn’t read right, we’ll have to rewrite it ... so it falls on us.”***

Consistency is critical because inaccurate, incomplete, or conflicting information on a Web site can be embarrassing at best, and at worst, lead to litigation. A public information officer considers the accuracy of information throughout the Web as the main benefit of using XML for content management: ***“The overall benefit would be the accuracy of information. And that’s very important for anyone, but certainly when you’re dealing with the customers that we have, that depend on the accuracy of information that we’re providing.”***

XML enhances consistency in two ways:

- XML’s ability to produce multiple HTML and PDF pages from single XML files. The management of thousands of individual Web pages is achieved via dozens of XML stylesheets (XSL) that produce those pages. This turns an overwhelming task into one that is highly manageable.
- XML’s single-source of content (described further in the next section) ensures that the same information appears the same across a Web site because it all comes from the same source.

Reduced data duplication and content handling

Data duplication means that multiple source files exist for identical text, images, and other content that appear in multiple locations and formats across a Web site and other media. This requires that the “same” content be handled in several locations, probably by different people, and often at different times. The absence of a single source file:

- creates greater risk of “version differences” between the duplicates,
- requires manual tracking of all locations of the duplicated data, and
- demands different technical skills for handling the data in its various guises (e.g., Word, HTML, database, etc.).

If content is modified in one instance, it needs to be modified in all its other occurrences, which can be an imposing task. One technical staff member characterized the process: ***“It’s sometimes difficult to get content up in a timely manner, again because of the multiple formats ... So there’s a big emphasis in the formatted Web pages. Then of course you have to do the full HTML document for accessibility standards. Then you have to do the PDF to actually get the full document for that. And as everybody knows, you need to make one change in one document while the other two things might not necessarily need a change. So you get multiple versions floating around [out] there all the time.”*** As Web sites grow, it becomes virtually impossible for any one person to remember where all of the entries are (see Figure 3).

Figure 3. Workflow in Non-XML Based Web Site

The diagram above shows how much activity is spent passing documents back and forth while trying to keep them all consistent and up-to-date. Many of the tasks consist of manual reformatting of the content for the Web, while checking that it's still accurate.

XML can eliminate the duplication of data because the XML file serves as the single source of the content. Its various manifestations throughout a Web site and beyond (HTML, Word, PDF, etc.) are produced via the XML stylesheets (XSL) that transform and present the XML content in the format and location desired, without modifying or duplicating the original XML source. XSL “handles” the content and produces the output, which not only eliminates the duplication of data, but also manages how that data is handled (see Figure 4 on next page). As one IT professional stated ***“Right now we have different Web pages for different types of documents like PDFs and different print-friendly forms and things along those lines. We have to change them in two or three places. So what we’ll do now is we’ll have the one document that will be accessed and we’ll only need to change it in that one place. So that will make for a much better environment and less work.”***

Compatibility with multiple devices and formats

The Web is relatively young and the surrounding technology advances at incredible speeds. Devices barely imagined in the early days of the Web (PDA’s, cellphones, iPods, etc.) are becoming commonplace; additional devices appear every year. As a technical lead on one of the XML Testbed teams said, ***“We’re going to see more PDAs, more personal, smaller, wireless applications that everybody’s going to want to deliver content to.”*** As a result, Web designers now must plan for more than basic desktop delivery, and content owners must envision their information disseminated across a broad spectrum of devices.

As these new technologies proliferate in the marketplace, they bring the compatibility, standards, and compliance issues that all new technologies bring. Web sites will need to adapt to support this new environment or, more accurately, *environments*. One Testbed participant emphasized the impacts that are already being seen: ***“Our legal staff and public information officer use BlackBerrys; other staff [members] use Palm Pilots and laptops, and a few others use cellphones ... making these types of formats available seems like it would be much easier with XML.”***

Figure 4. Workflow in XML Based Web Site

The diagram above shows how the single XML source document at the center of the process eliminates much of the redundancy and checking activity associated with the non-XML based workflow. Many of the manual tasks are automated.

It’s not that HTML-based approaches (including those that use dynamic scripting and database utilities) cannot handle multiple formats; they are just not designed for it. ***“Right now we have different Web pages for different types of documents like PDFs and different print-friendly forms and things along those lines,”*** remarked one Testbed participant on his agency’s current Web site management process. When changes occur to the content, ***“we have to change pages in two or three places.”*** Another Testbed team member described an alternative approach using XML whereby ***“the generating of the PDF and the Web page could all be done behind the scenes and on the fly ... just click the button, fix it, save it and then the print version’s updated and the Web version’s updated.”***

XML holds a big advantage over HTML in this regard because XML is a content specification standard (a meta-language of rules for how data can and should be described). Unlike HTML, it is not tied to an output format such as producing pages on a Web browser. Because XML is an open standard, it can easily adapt and integrate with new devices and formats. In the simplest sense, it only requires an XSL file to format the output to a particular device. And when content changes in the XML files, the XSL file immediately and automatically brings those changes to all the desired formats and devices. As summed up by one end user on a Testbed team: ***“Reusability—in terms of taking one XML document and being able to put it out in different formats and devices—that would be a big improvement.”***

Better information for Web site users

Using XML results in better Web site information from a variety of perspectives. The Webmaster benefits from the reduced effort required; users benefit from the more responsive service; and content providers benefit from the accuracy and timeliness of the information provided to the Web. As the Testbed participants discovered, XML offers a strategic advantage in this regard. One participant said, ***“It’s really the case that this is a forward-looking strategy.”***

Since the Web has become the primary vehicle for organizations to get information to their users, the challenge is to provide as much information as possible, in ways that are most useful to those users. From a business and public service sense, it is important that the information be timely, accurate, and effective. It not only demonstrates professionalism and competence, but mitigates potential bad will or lawsuits. XML can aid in this strategy because it dramatically reduces the time required for maintenance of Web pages (due to enhanced consistency and reduced duplication), while eliminating error-prone and redundant tasks in the workflow.

The highly automated framework that XML brings to Web site management increases confidence in the accuracy of the site while freeing up staff to produce higher value products for the Web. Testbed members found this cascading benefit in their own projects: ***“I think the biggest advantage you’re going to have is freeing up a really talented person to do more complicated work than is being done right now.” Or as one participant remarked, “Our Web site is getting exactly what it was getting before, except a little more and a little better, and it’s cost us nothing and it’s requiring no time, really, and it’s saving hours every day.”***

As Web sites continue to grow in importance, the public continues to become more savvy and demanding and increasingly expects high levels of service from them. When service does not live up to those expectations, the threat of alienating or losing these users increases. Because an XML-based Web site offers the opportunity to shift many of the time-consuming maintenance tasks to activities that improve the quality and responsiveness of the Web site, it can produce more consumer-oriented benefits. As one technical staff member said in regard to his project, ***“I don’t really think there’s a lot of resistance because everybody sees that it just opens a new avenue, because there are so many people out there that we really aren’t reaching, or we’re not reaching to the full extent. So by doing this project, it’s going to allow us to get those people in here.”***

Stronger foundation for data sharing and archiving

Data sharing, collaboration, and integration are dominant topics in today’s IT world. Organizations need to share data within their own organization and across organizations throughout the world. In addition, the **shelf life** of data is an increasing concern, especially as technology advances and formats once thought to be **universal** are now **obsolete**. As Tim Bray, co-inventor of XML and director of Web technologies at Sun Microsystems, stated at the XML Testbed Symposium, ***“XML is the best tool for creating a file format to ensure that things written today will have an excellent chance of being available for centuries to come.”***

The costs of developing and maintaining interfaces and middleware to communicate data across different formats can be prohibitive and shortsighted. It is far more advisable to use data formats that are open, standard, easily communicable and persist over time. XML is first and foremost an open, standards-based, data formatting specification. By its very nature, it is designed to enable the sharing of information because it is not tied to any device, technology, or proprietary software. By using XML—especially by adopting industry-wide standards within XML such as DocBook, EAD, and other data definition schemas used by the Testbed teams—organizations are building the elements of a shared information structure.

And the issue extends beyond data sharing to data ownership and accessibility. As Tim Bray also stressed at the symposium, XML provides organizations with the greatest assurance of content ***“longevity, reusability, internationalization, and vendor-independence.”*** In regard to ongoing access and archiving of that content, which is not captive to specific software or hardware requirements, XML offers the best solution.

Cost-efficiency in Web site and content management

HTML-based Web sites often require menial, repetitive maintenance tasks (checking pages for consistency, making the same changes in several different places, etc.), while XML eliminates most of them through its single-source, multiple output design. An agency staff member stated, ***“Its pretty straightforward to make conversions in XML documents quicker (than traditional methods) and more standardized so that there’s less wasted resources.”***

With HTML, cost efficiencies are inversely tied to the size of the Web site. It can be very cost-efficient to maintain a small site in HTML; but as the site grows, those efficiencies decrease with more pages and duplications of content to manage. With XML, the opposite occurs. Since the multiple pages of a Web site are generated by a very small number of XSL files, the number of files to manage stays constant as the occurrence of individual Web pages increases. For instance, an XML-based site with 20 XSL files may produce 100, 1,000, or 10,000 HTML Web pages. Regardless of the number of Web pages, the content still comes from single-source XML files, and those 20 XSL files produce all the pages. It’s a much easier management structure. (See Figure 5.)

Figure 5. Return on Investment for CTG in Converting to an XML-based Web site

As an IT manager from a large state agency clearly stated, ***“In terms of us actually doing the management of it [the Web site], I don’t see any problems. I can’t see where it’s going to do anything but save us time and resources, which mean money.”*** Likewise an individual serving as a technical liaison agreed on this benefit and linked it to XML’s single source capability: ***“I think just the notion that you’re creating that single source, which is incredibly important, you’re saving so much—you’re saving time, you’re saving money.”***

It is also important that with XML, staff time is not consumed by menial, repetitious tasks, but rather in work that will make the Web site more timely, accurate, and cost-efficient. A program staff member who works closely with the IT unit summarized it this way: ***“Well, the most important benefits I think would actually be sort of secondary benefits ... freeing up the Web unit from spending all their time creating HTML pages and altering HTML and tweaking stuff for people ... Having them freed up to do the more involved projects that we’d rather have them working on, would be a benefit for everybody.”***

Better coordination of publications and information

Publications present particular difficulties to Web sites due to their number of pages, unique formatting and layout, and navigation/paging requirements. In addition, most publications are created and maintained in a format that is “foreign” to HTML, such as word processing or desktop publishing software. Things that are taken for granted in many publications such as a table of contents, tables, graphics, and footnotes can be very difficult to recreate in HTML pages. Likewise, a single publication may have many incarnations on its way to the Web—from a word processing document (the “original”) to a desktop published document (the “printer’s original”) to a series of individual HTML pages (the “Web original”) to a PDF file (on the Web and in print). As a technology manager explained, ***“I actually happen to have somewhat of an example of that going on right now, this consolidated plan, this three-hundred page plan. They want to put it out—they had the version out there in PDF that was for public comment. Now they’ve gotten the approved plan ... And the question from the deputy for policy [was] ... what do I do [and] what do you want it in, what format? ... and I said, well, you need to get a PDF of it and we can put it out in PDF. If you wanted an HTML, you need to send us the Word document. The PDF can go out almost immediately once you’ve signed that this is ready to be posted. I said the HTML could take a week.”***

XML/XSL provides perhaps its biggest benefits in its ability to better coordinate publications. Since all the content for a publication can be contained in one single-source XML document, the problem with various versions and formats of the “originals” can be alleviated (see Figure 6 on previous page). Likewise, the peculiar challenges posed by publications for a Web page, such as the table of contents and footnotes mentioned above, can be “programmed” into a single XSL file and then applied to all the publications encountered on the Web site.

Figure 6. Creating and Maintaining HTML Web Pages via XML/XSL Files

In addition, one of the biggest challenges in the publication process occurs within the workflow. In most publication processes, once the document leaves the content developer and is handed off for review and edit, control of the source document can be compromised. In addition, different actors within the process can perform various jobs, so consistency and integrity can be compromised. A program staff member from a large agency explained it this way: ***“There are bottlenecks [in the] process, whether it is a piece of paper or electronic, ... it’s got to go through all those hands. The nice thing [about using XML] would be that ... we would just give them their piece ... to review and they could say, fine, and move on. So in that way hopefully it would make things move a little quicker.”***

Accessibility

A key principle of Web accessibility is designing Web sites and software that meet different user needs, preferences, and situations. Section 508 of the Rehabilitation Act of 1973 and NYS Policy P04-002 require Web sites to be accessible to persons with disabilities. ***“The process can be very labor-intensive bringing thousands of non-compliant HTML pages into compliance, but making it accessible might be a little easier for the Web unit, using XML,”*** said a technical Testbed participant. Properly structuring the data and style with XML can ease that burden since Web pages are generated automatically and uniformly. A change in one file can bring dozens or hundreds or even thousands of Web pages into compliance.

Furthermore, because XML separates content from style, it enables easier adaptation to new formats and

requirements that occur in the future. One Testbed Webmaster expects XML to help them ***“better meet the accessibility standard with properly-structured code and more flexibility ... rather than it was coded to do this certain thing a couple of years ago and now you have to recode it to do this new thing this year.”***
