Developing, Converting, and Maintaining Information-rich Resources on the World Wide Web

Mary M. Peet

SUMMARY. *Sustainable Practices for Vegetable Production in the South*, 174 pages long and with 250 references, was written as a traditional college textbook, but is also available as a World Wide Web (Web) site (<http://www2.ncsu.edu/sustainable/>). This article chronicles the conversion of the entire text to a Web document and the simultaneous release of Web and print versions. I will also discuss some of the issues that we will confront if we depend on the Web for delivering and receiving content-rich information. These issues are as follows. 1) Although there are no standards for Web sites as there are for print documents, there are certain similarities in the way most Web sites function. Relative to our familiarity with book and journal conventions, those of us educated in the age of print are unaware of Web standards. 2) The optimal size and structure of the information chunk is unclear. Should it be a whole chapter or article, a single paragraph, or a functional unit of facts that doesn't have a name or correspond to anything in print media? 3) Organization and consistency are critical. Table and chapter numbers are meaningless. The most important question is "How does a person accessing part of your Web site know about all the other parts and how they fit together?" You can flip through a book to view it, but a person following a link to a particular page on your site is like the blind man touching the elephant's trunk—the whole is hard to visualize. 4) There is no good place to put references and footnotes because of the subdivision of information into chunks of functional facts. 5) There is no obvious starting or stopping point in making revisions. 6) People accessing the site will send messages and ask questions.

I started this project with an unpublished book *Sustainable Practices for Vegetable Production in the South*, which I had been reproducing each fall as a course pack. The material was originally compiled with the help of a research assistant, Sarah Slover, whose work was funded by a 1991 SR-SARE (Southern Region Sustainable Agriculture Research and Education, then Low-Input Sustainable Agriculture) grant. Many faculty and extension agents at North Carolina State University (NC State) and elsewhere helped in developing and reviewing the material. In 1995, the Sustainable Agriculture Group on campus provided funds for copy editing, but I had not been able to find money to publish the material or a publisher willing to back such a specialized regional publication. The Web was an inexpensive way to reach a large audience, which included our county extension agents. As it happened, once the material was available on the Web, I was able to find a publisher, and the book version was published in Fall 1996. In fact, the publisher was able to download tables directly off the Web, a process that actually worked better than translating them from the WordPerfect 5 version.

In 1995, when work was begun on the Web site, however, we assumed that only the Web version would be available and we wanted to put the book online in a format easily accessible to a wide audience. The first decision was to use a hypertext format rather than Adobe Acrobat. Adobe Acrobat is a quicker and less labor-intensive way to put material online, but hypertext is more interactive and allows the reader to browse large documents and download only material of interest. Hypertext is also a good format for images of crops and pests and for video clips of farm machinery. The main advantage of hypertext, however, is the ability to link chapters to each other and to other Web sites. The second decision was to use Netscape as a browser, mainly because of its wide use on campus. A link was added, however, to a version of the text that does not require a Netscape browser for tables (low graphics version).

The third and most difficult decision was the size of the information chunks to include on a single page of the site. Chapters seemed too long, but individual paragraphs or subsections too short. Finally, each chapter was subdivided into three to four sections, plus large tables. In a few cases, the sections were also subdivided. An average Web page represents two- to three printed pages and about the same number of screens on most monitors. It is fairly easy to print out or to read just the information you are interested in, but it is time-consuming to print out or read the entire document.

The document was marked up in HTML (Hypertext Mark-up Language) in Fall 1995 and was mounted on the university server in January 1996. Since that time, problems have been corrected, more images have been scanned from slides, and internal and external links and a search script have been added. Use started to increase in March 1996, and for the week of 26 Apr. 1998 was 7336 hits. Usage statistics for this week are available as a link from the homepage. Obtaining and interpreting more detailed records of use of your particular site from university servers is usually possible, but not always convenient.

Project features

Why should someone use the site? Compared to other sites reached by surfing the net, the unique feature of

this site is that it is content-rich. The user is free to read, download, or print the equivalent of a 174-page book. Although the material is copyrighted, use is free with attribution to the source. The site received the ASHS 1996 Extension Materials Award for the Commercial Fruit, Vegetable, and Herb Production category. In 1998 it also received a SR-ASHS Extension Publication Award.

The homepage consists of three chapters related to sustainable production techniques (Soil Management, Cover Crops, and Living Mulches and Conservation Tillage); five chapters relating to integrated pest management [Introduction, Insect Management, Nematode Management, Disease Management (including a Disease Rotation Guide), and Weed Management] and twelve crop chapters (Beans, Cucumber, Pepper, Sweet Corn, Eggplant, Muskmelon and Watermelon, Tomato, Okra, Potato, Sweetpotato, Squash, Gourd, and Pumpkin, and Cabbage, Broccoli, and other Cole Crops). Chapters are subdivided by topics. All the crop chapters are subdivided into Botany, Production Practices, and Harvest and Postharvest. Long tables are linked separately, but short ones are included in the text. Additional choices on the homepage Table of Contents are the Preface, which describes how and why I wrote the book, Economics (budgets for crops grown organically in California), a description of calculating degree days, a list of organic certification programs in the South, and resources, including useful publications and addresses of cooperative extension offices throughout the South. An additional section, not present in the textbook version, is Other Web Resources, which provides hotlinks to Web sites at NC State and throughout the world. There is also a keyword search feature.

Comparing book and Web site features

Should someone who needs information about sustainable vegetable production visit the Web site or read the book? Anyone who wants to learn the source of a particular piece of information will need to use the book. The online version lists all references, but does not footnote specific information. Books are more convenient to read and to take on the road, but the Web site allows printouts and connec-

tions to specialized databases. To locate specific information, book owners can use the index, while Web browsers can use the search feature. I think most people who have been raised using books will find it easier to orient themselves on the material and to find specific information using the book, but the next generation may prefer the Web format!

Development issues

WHY WOULD YOU WANT TO PUT MATERIAL ON THE WEB? Even commercial Web sites have difficulty making money on the Web. I am sometimes asked why I developed a Web site when it could compete with book sales and "You don't get anything out of it." In what sense can a strictly educational site be profitable? I was faced with having material in hand, but no publishing outlet. However, for HortBase to work, people will have to share their work more or less on faith that Web publishing is worth the time, money, and effort. Our universities support Web site development in general and provide servers and some resources, but I'm not sure they are set up to reward people who put a lot of effort into Web site development. Current reporting practices don't distinguish between putting a syllabus or resume online and putting a text or major resource online and maintaining it. I think one of our priorities as a group interested in Web publication is to work with our institutions on recognition and incentives for developing good Web sites. To provide incentives for people to develop and share resources, we need to think about how to recognize Web resources. The ASHS Extension Materials award this site received in 1996 is one example of a way to recognize and encourage Web site development.

For me, however, the main reward has been the use of the site. In one week the site was accessed by 778 individuals or workstations, 20% of which were from outside the United States. This compares to <1000 copies sold of the book! Believe it or not, I actually enjoy hearing from people all over the world who write in saying they like the Web site even if they also ask more questions! I have heard from people who have used my material either in or for planning projects for Taiwan, China, Ghana, South Africa, New Zealand, Australia, and even a greenhouse project at 14,000 feet in Tibet!

The best reason for Web docu-

ments is that this will be the way material is communicated in the future. The Internet has doubled in size from 6.6 million hosts (a domain name that has an IP address, basically representing a computer connected to the Net) in mid-1995 to 12.8 million hosts in mid-1996. In the same period, the number of domains (such as ncsu.edu) went from 120,000 to 488,000. It is harder to estimate the number of users. Nielson calculated 16.4 million U.S. Internet users in 1995. Worldwide, the number of users has been estimated to be 23.5 to 26.4 million (Kantor and Neubarth, 1996). Institutions like Texas A&M and Ohio State, which have been at the forefront of developing good information and links, will continue to be seen as the best sources of horticultural information. An institution without a significant Web presence will become irrelevant. Just as one example, even though there is no institutional affiliation on my site, I have been contacted by potential graduate and undergraduate students and postdocs who saw the Web site. No school can afford to ignore online shopping for educational opportunities.

THE STARTING POINT. The best and worst feature of Web site development is that there is no clear beginning or end. Material can easily be added chapter by chapter, unlike a printed publication, which goes unread until finished. You can get a quick start by putting up an outline and then filling in material later. The problem with an incremental approach to Web development is that, when a new Web site comes online, there is a flurry of activity as the Web search engines add it to their indexes, and it is often featured as new. If the site is disappointing, search services may not pick it up or, if they do, people may not bookmark it, link to it, or pass the word to others. I think there is less tolerance now for sites that are under construction because there are so many good sites. Another problem in going online too early, is that the search services may not index everything. Of course some services automatically search the entire Web, so in that case it wouldn't matter.

At most universities you can find students with HTML experience who will work for \$7 to 8/h—much less than commercial developers charge. Since it's a new technology that looks good on resumes, students may be

willing to construct a Web site for special project credit. You can also get consultants fairly cheaply on campus. If someone is setting up the site or server, they may be willing to take your text and mark it up for you and mount it on their server just to increase the material offered on their site. They may not want to add a lot of features, such as search engines or links, but at least the text would be available.

THE SIZE OF THE INFORMATION CHUNKS. The size of the information bite is another important decision. In a book, the division into chapters, subsections, paragraphs, and sentences is something everybody understands. No one really has a handle on the optimal size for a Web information unit. Ideally, the information chunk should be free standing, i.e., understandable and indexable on its own. Ideally, it should also be viewable without too much scrolling. Of course this depends on monitors and browsers, but probably comes down to something shorter than a book chapter but longer than a paragraph—a chapter subsection. I like to think of this unit, as a functional fact, as opposed to a factoid, which contains so little information that it is interesting only as trivia. Most of the Sustainable Practices pages can be scrolled through in two to three screens and printed out on one to two pages of legal-sized paper. Even this level of subdivision required significant regrouping within the book chapters, however.

HOW SHOULD IT BE ORGANIZED? This is one of the most difficult, but at the same time critical, questions. There are lots of how-to books available on the topic of Web creation and organization. Most Web sites grow haphazardly rather than reflecting elegant and information-efficient designs. Sustainable Practices follows the book Table of Contents fairly closely; but, if I were starting over to create a Web document, this is probably not the format I would choose. Unfortunately, it is hard to change the organization of a Web site once it is created without breaking links other people have created to your site—another good reason for waiting to mount the site. The average life span of a URL is only 45 days, and it seems to me we have a responsibility not to add unstable links.

An important aspect of this organization is consistency within the Web site. A reviewer pointed out that the way titles were listed in the homepage Table of Contents was slightly differ-

ent from the way they were listed on the link itself. While not confusing in a bound print document, in the ill-defined boundaries of hyperspace, a user might be confused about whether he or she had linked to the correct place.

Another early issue was dealing with numbering tables and chapters, which is actually a process of denumeralizing the entire document. A linear sequence of numbers isn't very meaningful in a Web document since there is no reason to access the document in any particular order. The chapter numbering issue was relatively easy to resolve by creating links to chapter titles that were fairly short. Tables don't really have titles, and captions were generally too long to use, so short descriptive titles had to be invented for a link reference in the text. Then these link references had to be incorporated into the text. Rather than repetitive statements of "Click here for information on...", we tried to integrate the table reference smoothly into the narrative. This chapter and table denumeralization made it more difficult to cross-reference between the book and Web site, however.

ASK THE EXPERTS. If possible, the final step before mounting the document and every 6 months thereafter should be to have someone with Web experience inspect the site. Web conventions change rapidly, with new Netscape features being added all the time. In a recent article on a hot Web site developer of push technology, a Time magazine writer referred to a generation on the Web as lasting 6 months. Andrew Kantor (1997), a senior editor at Internet World, described a lifetime on the Web as a few years. A site that isn't given a facelift every so often will tend to look dated, like a house with an avocado green refrigerator and harvest gold shag rug! Features that I suggest avoiding include frames, animated applets, banners, counters, and brightly colored backgrounds and textures. These are very distracting if not done well and slow up downloading on older machines or those with slow connections. Push technology also now seems passé. NC State has been very active in promoting Web-based teaching. As a consequence, I have been able to obtain expert critiques of my Website and help in implementing the suggested improvements from staff in the newly created Learning Technologies Cen-

ter. Currently, for example, someone from the center is working with me to integrate video clips of farm equipment into appropriate chapters.

OTHER CONSIDERATIONS IN WEB SITE

ORGANIZATION. Most of the ways we are used to getting information have an enforced linearity. Books, lectures, movies, television, and radio broadcasts are designed to have a starting and ending point. A Web document gives the viewer freedom to jump in and out of the information flow at any point, but at the same time forces the Web developer to structure the user interface so that the groundwork is laid for any concepts that are introduced. It will be interesting to see how this issue is resolved given the rush to adopt distance learning and the rising impetus for Web-based education. It seems to me a whole new educational pedagogy will need to develop before we get very adept at this. Those of us locked in the print age by our early education may never develop the skills to teach very well on the Web. However, as people who are developing information resources, such as HortBase, our mission is considerably less complex. We can probably safely assume that readers are already interested in the subject and understand the basics but need specific facts, which we, the content experts, are in a unique position to supply. We need only provide a logical, easy to follow format and good information.

At the same time, the big advantage to putting material on the Web in a hypertext format is that you can organize the chapters several different ways—the traditional table of contents and some other pattern. I found it very valuable to reinforce a systems approach by inserting links between chapters. A criticism of the book had been that the crop chapters included little material on cover crops, IPM, organic additions, or other practices that people associated with sustainable production. This was less of an issue in the Web document, because it was so much easier and less awkward to cross-reference. For example, if a rotation study discussed in an IPM chapter includes corn and tomatoes, I could insert a link into and from the corn and tomato chapters. This brings up the matter of how to weave links seamlessly into the text. Integrating material between chapters was difficult and involved re-writing and further drift apart of the

book and Web site versions to avoid replacing “See Chapter 1 for further information on ...” with lots of sentences saying (“Click here for information on. . .”). I avoided direct links from my text material to external Web sites because of the difficulty of checking for and replacing broken links scattered throughout the entire Web site and forming an integral part of the text.

HOMEPAGE DESIGN. Features you want to highlight on the homepage are another important question. The homepage design should be inclusive and draw the visitor in but not be cluttered. Imagemaps are currently a popular way to do this and are very efficient in the use of space and in describing relationships between areas. Good homepage design should highlight themes, special tables, newly added material, or other features. Consistency with your university or departmental homepage in terms of logo, color scheme, or navigation system is also a nice feature. We may want to consider standardizing the look and feel of HortBase to give the viewer a sense of connectedness of the entire document.

The main limitation to taking advantage of hypertext features is our own linear mindset. For example, I would like to do a major update to the information on my book to include recent research findings. In the book, I will probably insert sentences into the text flow, a time-consuming job and possibly one that is distracting to the central theme. I feel that the much-edited current document delineates the basic information succinctly and well, and I am reluctant to insert a lot of detailed information from research reports. In many cases, research reports require interpretation for nonscientists and it is not clear how applicable they are to other geographic regions and cropping systems. An idea I am currently considering for the Web site is to put all the updates together as an appendix, which could be highlighted on the homepage and linked to relevant text material in the appropriate chapters. With this approach, people who are revisiting the site can easily view the new material, but not be distracted from the main themes.

Navigation aids are very important in a large Web site. In a book it is always clear that a chapter or paragraph is part of a whole, and it is clear

how the size compares and other functional relationships. There is no bound-together volume in a Web site to orient the reader on the total size of the document and how the separate pages relate to each other. What I didn't realize in initially developing the site design is that people would connect either through a search or someone else's link to just one chapter or table may never reach the homepage or realize that the page they are reading is part of a large Web site. While it is important to have each Web page self-contained, it is also important to have navigation aids, so that people can move to related topics or the immediately previous or following Web pages. Navigation aids should at least link to the homepage. A navigation bar on each page to get to the other chapters would be nice, but would take up lots of space if it included meaningful chapter titles, since chapter numbers would be meaningless. Constructing imagemaps to graphically convey relationships is the most space-economical, though challenging, solution.

HOW FANCY DO YOU WANT TO MAKE THE SITE? The answer to this depends on finances, your expertise, and the expected clientele. When I started my Web page in 1995, I was warned against putting a lot of graphics on the pages because some people didn't have graphic browsers. I don't think this is much of an issue in 1998, but it is still a good idea to minimize pictures or put them in special galleries or as thumbnails to speed downloading. Real audio and streaming video are also problematic because so many people have slow modem connections just want to take a quick look. Although I originally wanted to insert audio and video segments, they required disproportionate equipment, expense, and expertise compared to text, images, and links.

COPYRIGHT ISSUES: YOURS AND THEIRS. Having your name, the name of the Web site, and a link to the homepage on each individual page will provide some identification for your material if it is downloaded or printed out. Some people are very concerned that their material will be easy to copy and they will lose their copyright. All material is protected by copyright laws and putting material on the Web does not compromise your rights as an author in any way. Students downloading a chapter and turning it in as a term

paper are as guilty of plagiarism as if they had copied it by hand out of a book! Unfortunately, plagiarism is easier on the Web, because material doesn't even need to be retyped. Hopefully most people will act like the Aldo Leopold Center student, who used my material in her term paper and wrote to thank me because she got an A! Colleagues using our Web material and images may not know how or whether to cite a Web site, however. It seems to me that ASHS and this working group could set standards for using Web material and publicize the standards that have evolved for citing Web sites in publications and linking rather than copying from other Web sites.

In terms of violating others' copyright on your site, the issues get fairly complex if you are using material that is not in the public domain. With the current uncertainty in fair-use legislation, most people are being cautious. In most cases, the issue is probably professional ethics and academic integrity, just as in print documents. As long as we give credit to other people for their contributions, I don't think the issue of copyright will be a major limitation for academics. One area I personally find difficult to handle is keeping track of images in my slide collection, which I have copied from other people or which they have given me. I apologize if my Web site includes images from anyone else and I haven't credited your slide! I would also like to be able to use slides from the American Phytopathological Society, but I haven't succeeded in getting permission.

WHEN DO YOU STOP? When you run out of money you have to stop! I think everyone that has a site always wants to expand it in terms of adding new information, pictures, updates, and new features such as course materials, PowerPoint slide sets, etc. For most of us this is an add-on responsibility and it's unclear how much recognition we get from Web sites in terms of our professional responsibilities. A role for HortBase would be to coordinate, formalize, and legitimize these activities so we would have more recognition from our institutions.

MAINTENANCE IS AN ISSUE. Internet magazines suggests checking links at least monthly. The problem with updating is that unmounting and remounting the new material takes time and can also introduce errors. There

are scripts and software packages that do this (at least the link checking), but many server Webmasters restrict the scripts and software that can be used. Of course you can have your own server, but it's nice to have university security and backup. This may change as hard-drive space gets cheaper and site software more user-friendly. At this point most of us probably don't really want the headaches and expense of maintaining a server if someone else will do it for us.

DO YOU WANT TO BE MORE VISIBLE?

Although I don't actively solicit responses, I still get a lot of e-mail. Most people say nice things, some have questions I can answer or at least comment on, and some ask totally unrelated questions. I have forwarded some of these to colleagues. Usually, however, I don't pass on the colleague's e-mail address for fear they will reciprocate!

WHAT DO I WISH I HAD DONE DIFFERENTLY? Citation practices in hypertext have been the most difficult issue for me and have ended up as the part of the site least satisfactory to me. Interestingly, however, there have not been any complaints from visitors. In the original course-pack version of the text, references were numbered in order of citation in the text and listed at the end of the document. Numbers increased chronologically from page 1 to 174, although if a reference were cited a number of times the footnoted number would stay the same. Numbers were used rather than the Harvard style (e.g., Smith, 1998) to save space and make the book more readable for nonscientists. This method worked well as long as we stayed within one version of one word-processing package. However, when the material was copy edited, references were moved from the end of the book to the end of each chapter to make them more easily viewed by the reader. This involved some duplication, since references used in more than one chapter would have to be listed more than once, and we renumbered all the footnotes in each chapter. This was done, however, and is the current format for the book.

The problem in the Web version arose when the chapters were further subdivided. There was no logical place to put the footnotes without continuing to subdivide and renumber the references on each Web page—a tedious and error-prone process and one in which some references were repeated

many times. Finally we ended up putting all the references with the introductory material at the beginning of the chapter. We ran out of time and money at that point, and as a stopgap measure I hid the actual footnote numbers (although they can be viewed in the source code) so that people would see an unnumbered list of references at the beginning of each chapter but would not see specific footnotes in the text. On the homepage, I offer a link to a complete list of the references but refer the reader who wants specific citations to the book version.

No Web convention has yet evolved that I am aware of to handle this problem. Probably the best way is to put all the footnotes together as an alphabetical list. The way to handle the footnotes within the text is a more difficult decision. For scientific papers, the most elegant solution is to use the Harvard format in the text, with a hotlink to either an anchor by the name in the reference section or just link to the alphabetized page of references. I'm not sure what degree of documentation is necessary for general audiences. Numbered footnotes are hard to keep up with in a Web site if more references are added. Even in college textbooks, only table and figure references are specific. A list of selected references or a bibliography generally appears at the end of a chapter. Specific footnotes are most valuable for the person keeping track of the site so that the original material can be checked if a question arises.

At various points I felt I was really hampered in my thinking about the site because I was a book person in my overall orientation and not just because this Web site started life as a book. It's hard for me to imagine developing material that doesn't look like a book. If I had never seen a book, I could have made a better Web site. Paul Gister (1997), in his new book "Digital Literacy," suggests that the Internet is a fundamentally new medium requiring radically new approaches. He writes in the introduction "We're used to television and radio, two ways of communicating that work from the center out. Both call for a passive approach from their audience; we put ourselves in front of a receiver and absorb the content offered by networks and local stations. Where the Internet model diverges is that it places responsibility in the hands

of the individual. Rather than being spectators—information consumers—we become Internet users, people who discover and evaluate content before deciding how to put it to work.” It will be a continuing challenge for all of us to use this new media in ways that are creative yet educationally sound.

My ten best ideas for content-rich sites

- 1) Decide early on how the material will be organized, then divide it into chunks one to two screens long (or whatever unit seems most appropriate for the viewers you anticipate).
- 2) Add navigation aids on each page to the homepage and other parts of the site. Be sure titles of the sections are consistent between the link reference and the actual material.
- 3) Put your name, the title of the Web site, and your institutional affiliation at the top of each page where it will be printed out.
- 4) Try to recruit someone to help you with it who is likely to be available in the future. Retainers for monthly maintenance help if funding is available.
- 5) Many universities have a graphic standard for their servers. Consider a front-end for the site that uses these graphical conventions, images, and color themes.
- 6) Decide early how to handle footnotes, literature citations, and numbering of tables and chapters.
- 7) Keep graphics, backgrounds, and other Web features simple and unobtrusive. Downsize your expectations relative to video, audio, and other flashy features.
- 8) Once most of the material or at least a prototype is ready, ask people who have done a lot with Web development to make suggestions on organization and Web features.
- 9) Read (either online or in print) and attend as many seminars as possible on Web-related topics to get a feel for the issues and trends and to meet people who can help you.
- 10) Downsize your expectations relative to competing with commercial and entertainment Web sites then relax and enjoy your e-mail!

Information resources cited

- Glistler, P. 1997. Digital literacy. Wiley Computer Publ., New York.
- Kantor, A. and M. Neubarth. 1996. Off the charts. *Internet World* (December):46.
- Kantor, A. 1997. Microsoft vs. Netscape. Who will win, how long will it take, and does anyone really care? *Internet World* (January):116.