Use of the World Wide Web has grown to the point at which the Internet strains under its own popularity. Nevertheless, it is an increasingly valuable research tool for biologists (Smith 1993, Robison 1997). Powerful Web search engines allow users to locate material on virtually any topic, from pertinent literature to the whereabouts of colleagues. The Web offers an effective tool for project documentation and education, and the simplicity of making resources available via the Web encourages free exchange of data for peer review, collaboration, or public dissemination (e.g., Mars Pathfinder mission information, data, and images).

Researchers and educators who create and use biological Web sites are often the same scientists who write and read scholarly articles in printed publications; however, these two outlets differ beyond their media. Web pages (other than e-journals; Walker 1996) do not undergo formal review, they often have a larger readership than printed publications (see below), and they are indexed word-for-word for a worldwide audience. Nevertheless, Web pages are seldom prepared with the same scrutiny given printed publications. Authors of professional articles know that through peer and editorial review, their material will be subject to close examination and "gatekeeping" by colleagues. Web pages do not have to survive this process; they are easily published, appear in a medium lacking built-in prestige, and usually receive little critical review, even when user feedback is solicited. Moreover, Web sites are seldom, if ever, entirely finished.

All of these factors make it especially incumbent on those who put resources on the Web to exercise care in selecting and presenting the material. In this article, we describe our experience making material for research and scholarship available on the Web. Our site, which we established in early 1995 at the Illinois Natural History Survey, is used to distribute free software for wildlife ecology and related disciplines (Diehl and Larkin 1996). We also discuss data we collected about the use of the site. Our article is not, however, a recipe for establishing a Web site; different operating systems and a proliferation of Web authoring tools make such recipes useless for a general publication.

Splashy features versus thoughtful prose

The Web represents a medium with features that can be used to create a pleasing, intuitive, and sometimes interactive presentation, but these features should be used in moderation because they carry a cost: Splashy images, sound files, movies, and other features can require minutes to download, in contrast to seconds for formatted text. Such delays disproportionately affect those accessing the Internet from home, many of whom use a modem rather than a high-speed link (approximately 70% of users; Pitkow and Kehoe no date), and from geographically distant locations. Therefore, most Web pages should be designed so that voluminous graphics and similar features are optional rather than prefixes to essential text material.

Although download times have become an increasingly important issue driving the design of Web pages as increased use has strained the Internet to capacity, in some cases multimedia features are essential to the quality of Web presentations. Indeed, Web presentations require updating as the technology becomes more capable (Chinn and Bledsoe 1997).

Feedback from users and interaction with contributors

Seeing a site through the eyes of an outside user offers site maintainers an impartial perspective, but surprisingly few users of our site have contributed their feedback, despite our taking every opportunity to ask for comments. Ideally, users serve as online reviewers, notifying Web site maintainers of access problems and offering input on organization, presentation, and content of pages. Over time, user feedback becomes a tangible measure of interest by the Web browsing community and helps to provide an impetus to keep the site accessible and up-to-date.

In addition to users, some sites also serve contributors, those for whom the site offers an outlet for their substantive material (e.g., as in our case, new computer software or updates to existing packages). Contributors also rely on the site as a place to refer interested parties instead of maintaining their own site or copying and mailing floppy disks. Similarly, site maintainers rely on contributors for content and new material. In effect, a symbiotic relationship exists between a site's maintainers and contributors.

Material on Web sites should be accurate and available

Web addresses change and links to other resources expire, which has lead to a chronic infestation of "Not Found 404" errors. Web maintainers should alert themselves to the existence of such expired links by using link-checking programs (e.g., SoftSeek 1996). Similarly, dead-end

by Robert H. Diehl
and Ronald P. Larkin

April 1998
Figure 1. Cumulative number of Web and ftp site users over the first six months of service (25 January 1995 to 10 July 1995). Total users (top line) equals new (middle line) plus repeat (bottom line) users. Arrows mark announcements in the Internet newsgroups (N) sci.bio.ecology and sci.bio.conservation on 10 February (when curves begin to rise), then simultaneously in the Ornithological Newsletter (O) and sci.bio.ecology on 4 April, and finally in the Wildlife newsletter (W) on 30 May.

links to “under construction” pages waste users’ time and serve as an announcement that the site is not well maintained.

In addition to requiring the same regular backup policies as other computer resources, networked computers— including Web sites—are increasingly subject to deliberate attempts to bring them down or compromise their availability to users. Having experienced such break-ins, we can say that they are traumatic. The security of a computer system is compromised, the Web site is unavailable; it fails its users while they are surprised to find that such intrusions more frequently come from within an organization than from an unknown operator sitting at a distant computer (Garfinkel and Spafford 1996, CERT 1997).

Maintaining a Web site that serves as a software archive requires additional effort. Such software is usually made available via an ftp (file transfer protocol) server. Contributors can upload their software directly to the ftp server rather than sending floppy disks through postal mail or e-mailing their software as large attachments. We assess the appropriateness of uploaded software and other material submitted by contributors or copied (with permission) from other online sources. This step also serves in verifying that the software functions as the contributors describe. Before adding new software to the site, we scan it for viruses using reputable, regularly updated virus-checking programs. New software may also require compressing or archiving that is particular to the target computer— operating system combination. Because different software contributions are designed for different computer systems, these operations require having access to multiple computer platforms and operating systems, as well as the technical expertise to work in those environments.

When the security of a computer system is compromised, the Web site is unavailable; it fails its users while the discomfited site managers painstakingly try to discover the extent of the damage and take steps to prevent another intrusion. Many site managers are surprised to find that such intrusions more frequently come from within an organization than from an unknown operator sitting at a distant computer (Garfinkel and Spafford 1996, CERT 1997).

A Web site lacking visibility serves no one

Unless a newly created Web page is linked to by other Web pages or has been publicized, users will remain unaware of the site. A site’s existence can be announced through judicious postings to relevant Internet newsgroups (e.g., in appropriate sci.bio newsgroups or in comp.infosystems.announce) and by submitting URLs to Web search engines (e.g., DEC 1996, Wired Digital 1996, Yahoo 1997). Each announcement of our site on Internet newsgroups or by submitting URLs to search engines as new users became aware of the site (Figure 1). The amplitude of this pulse decreased with successive announcements, perhaps because most interested parties who read these publications had been reached by the time the final announcement was published. After the first announcement, increasingly more site activity was maintained through links from other sites and by matches from Web search engine queries. These two sources account for nearly all site activity today. It is important to remember that interdisciplinary service efforts require interdisciplinary notification. Site developers should know that search engines use material in Web page titles, keyword lists, and the text body to index Web pages (DEC 1996, NCSA 1996, Wired Digital 1996).

Web site managers often monitor “hits” (a “hit” is a record of when and from which Internet address a page is accessed; we keep these records confidential) to justify the effort of developing the site and to motivate themselves to maintain it, but there are some problems with counting “hits” as a measure of Web site activity. For example, a single user often accesses a site more than once in a visit, and many “users” are actually automated Web robots gathering data on Web sites. Nevertheless, measuring hits is a reasonable way to estimate use as long as hits are not equated with actual human
For two and a half years, our moderately specialized site has been accessed by users in more than 70 countries: It is seen by over 30 different users a day, two-thirds of whom are first-time users of the site (Figure 1). This level of exposure, which is common among Web sites, likely exceeds that of most printed scholarly articles (Tenopir and King 1996).

Whereas conventionally published scientific articles are generally peer reviewed and reach a small, specialist audience, most documents and other material on the Web lack external quality control yet can have a huge audience. Announcing Web sites in scholarly newsletters (Figure 1), making the site available to a 24-hour planet, and using fancy features in moderation make it easy for people to use the site. Maintaining confidentiality and security, keeping links current, seeking feedback from users, and checking downloadable software for viruses are less obvious but necessary to serve those users.

Acknowledgments

We thank Jim Crowder, Paul Pomes, and Mary Bhuthimethee for providing hardware and services essential to our site's function. Additional thanks to Thomas Burk and three anonymous reviewers for helpful comments on an earlier draft of this manuscript.

References cited


Robert H. Diehl (e-mail: rdiehl@uiuc.edu) is a graduate student in ecology at the University of Illinois and the co-Web administrator at the Illinois Natural History Survey, Champaign, IL 61820. He studies avian ecology and technical applications to wildlife studies. Ronald P. Larkin (e-mail: r-larkin@uiuc.edu) is an associate professional scientist at the Illinois Natural History Survey, Champaign, IL 61820 and is also adjunct associate professor in the Department of Ecology, Ethology, and Evolution at the University of Illinois. He studies technically oriented field biology including radar ornithology, radio tracking, real-time data-acquisition, and automated detection and monitoring of animal sounds. © 1998 American Institute of Biological Sciences.