

ISSN 0701-3086

PRE-PUBLICATION PAPER

Commercial Internet Publishing -  
the practicalities

John C. Nash  
and  
Mary M. Nash

WORKING PAPER  
96-39

July 1996

**This working paper should not be quoted or reproduced without the written consent of the author.**

## Commercial Internet Publishing - the practicalities

John C. Nash, Faculty of Administration, University of Ottawa,  
136 Jean-Jacques Lussier Private, Ottawa, Ontario, K1N 6N5  
fax: (613) 564 6518 email: jcnash@uottawa.ca

Mary M. Nash, Nash Information Services Inc.,  
1975 Bel-Air Drive, Ottawa, Ontario, K2C 0X1  
fax: (613) 225 6553 email: mnash@synapse.net

### Abstract

This paper is an update of our 1995 presentation, *Internet Book Publishing -- a case study*, where we reported our experience in publishing *Scientific Computing with PCs* using File Transfer Protocol (FTP) under a shareware model.

While this experience was useful, we believe "begware" to be an impractical mechanism for commercially viable publishing. The present discussion will consider the costs of Internet publishing, approaches to its implementation, decisions regarding the packaging of products, marketing strategies, pricing, and possible payment schemes. Our views are based on efforts in progress to publish several specialty works.

This paper was published in *Electronic Publishing: Its Impact on Publishing, Education and Reading* (ed. Charles T. Meadow, Maggie Weaver, Françoise Hébert), Canadian Association for Information Science, Proceedings of the 24th Annual Conference, Faculty of Information Studies, University of Toronto, pp. 62-72, 1996

## Introduction

At the 1995 CAIS Annual Meeting, we discussed the Internet publishing of our book *Scientific Computing with PCs* (Nash and Nash, 1995), which we abbreviate to *SCPC*. Our purpose was to experiment with the Internet as a mechanism for the publication of works that have been, until now, conventionally published on paper by a batch printing process with mail-order or bookstore distribution using direct mail and print-media advertising. This paper is an attempt to expand the ideas of that exercise.

First we consider some of the issues identified in the *SCPC* venture. We look at ways that have been or are being developed to overcome some of the difficulties or to take advantage of particular features of the Internet for publishing.

While some commentators (see for example, Okerson and O'Donnell 1995; Peters, 1995) have suggested that publishing on the Internet will be considerably more than a different or alternate channel for content, we do not explore such ideas here. That is, we restrict our attention for now to the topic of marketing and distribution of intellectual property that would usually appear in print form. We have, however, started preparing content in the form of hypertexts with embedded software, graphics, sound or video files.

The content we wish to publish via the Internet, in addition to *SCPC*, includes:

- the diaries of Arthur Slessor from 1888-1891, the diary of a Huguenot ancestor of one of the authors, and tapes of an interview with the late J H Wilkinson concerning the construction of the Pilot ACE computer by Turing. These and similar works we group under the project name *First Person History*.
- *Nonlinear Parameter Estimation: an Integrated System in BASIC* (John C. Nash and Mary Walker-Smith). The rights to the book and software were recently returned to the authors. *Software Extensions and Example Problems*, not released before except in private distribution, will accompany it.
- various commentaries on particular mathematical, statistical or computational problems, of which the file `Q Q J N 9 6 0 1 . H T M` is a prototype, albeit freely available. (URL=<http://macnash.admin.uottawa.ca/QQJN9601.HTM>)
- a number of our articles still of interest to other workers that are exceedingly difficult to acquire, for example, Nash and Price (1979). By use of optical character recognition and image scanning we created versions of this paper that we were able to supply to a colleague who needed them for research in optometry. The request came to us while we were in the Netherlands and we were able to immediately send the files by electronic mail to the University of Waterloo.

These examples cover several styles of publications that may find customers, particularly on the Internet. We will now consider how such content may be "packaged" for delivery and use so that authors' and publishers' rights are protected yet the customer does not experience undue inconvenience. We shall also look at evolving possibilities for marketing and order fulfillment.

## Packaging of Internet products

The "packaging" of content for Internet publication parallels that in traditional publishing, or at least it should. Sadly, many Internet products lack proper writing and editing. The types of content that interest us here are largely text and graphics that were formerly on paper. A high standard of editing and page design and layout should be maintained:

- Spelling and grammar should be correct and follow acceptable conventions;

- The position and size of graphics should be pleasing;
- Fonts and font sizes should be clear and suitable to the material, and the use of italic and bold faces should be sparing and consistent;
- Headings should be sized appropriately and consistently;
- An index should be provided when appropriate;
- Suitable editorial notes should explain any unusual features.

Our intent is, of course, to transmit the material electronically to our customers for local printing. This raises some additional concerns.

We will not have control of the printer or display systems that our customers will use to read the documents. If many of our customers will read our product on a screen, then the fonts chosen should be large enough to make this workable, yet the lines must be structured so that screen display does not "chop up" our pages badly. European customers will generally have A4 paper (8.27 in. x 11.69 in., or 21 cm. x 29.7 cm.) and will not have access to legal sized (8.5 in. x 14 in.) paper. A little attention to detail can overcome the differences between North American letter size (8.5 in. x 11 in.) and A4. See <http://macnash.admin.uottawa.ca/dme00.htm> or Nash (1993).

We believe that users prefer documents printable in 2-up mode. That is, two pages of material are printed on a single letter-size sheet of paper. We have done this with both the re-publication of Nash and Walker-Smith and with the Slessor diaries.

Optical scanning and character recognition (OCR) introduce special concerns. Low-cost OCR has improved greatly in recent years. Documents having good contrast, adequate print density and fairly conventional font selections and sizes can be converted to word processing documents with few errors.

The Slessor diaries consisted of 280 Quarto (8 in. x 10 in.) sheets that had been typed on a portable manual typewriter. We still have this venerable machine. The diaries served as an OCR test at the University of Ottawa in 1993. These test files proved adequate for editing and did not need re-scanning.

Despite variation in ink density that attends manual typing and some (albeit careful) hand printed corrections, we found a very low rate of errors. Unrecognized characters occurred only 2552 times in a plain text version of the diaries that is 425,516 bytes long. This is a "not recognized" rate of less than 6 per 1000. We do not know how many times misrecognized characters occur, that is, an incorrect character is sensed, but we believe, from our editing, that it is lower than the "not recognized" rate. Note that the file length quoted above is slightly misleading, since it includes spaces, line feeds and carriage control characters.

Diagrams and photos present their own concerns for printing. The capabilities of printers to render fine detail, to a precise register, vary. Here we will only consider monochrome / grey-scale printing. Simple line drawings are straightforward to include and print. Photographs or drawings with many fine curved lines are more difficult, since ink or toner deposition must be quite precise to give a good rendering.

To avoid providing multiple versions of documents, we have so far chosen to provide documents in PostScript™ (PS) format. This is a plain text command language for printers that works consistently across makes of printers. GhostScript programs to display and print PS files on many computing platforms and printers are freeware, though the setup and configuration of these is tedious. A disadvantage of PostScript, however, is its very slow rendering of graphics, especially photographs. To date, this has not concerned us greatly, since our publications have used primarily line drawings; it will become more important as we widen our product line.

PostScript is unfortunately **not** a format with which a large proportion of general users of the Internet are familiar. On PostScript printers the file format is hidden from view; sending PostScript files to a non-PostScript printer usually wastes a lot of paper and ink. On the other hand, it is relatively awkward, though not impossible, to convert

PostScript to machine readable content suitable for word processing software. We have developed programs to put a header or footer on each page that identifies the copyright holder, thus making the material less attractive to those who would, by simple editing or photocopying, resell our products as their own. We are not worried if a single user makes a second copy so one can be marked up for working purposes. What we do not want is the wholesaling of our materials, since we anticipate most of our products will have a widely dispersed market with relatively small total sales.

Our ideal in "packaging" is to provide materials in a form that the customer finds attractive and easy to use but that, at the same time, carries reminders that the products are copyright and require payments to the publisher if copied. From an operational perspective, we want to have tidy products. This means that we want them to reside in single computer files, even if these are archive files. As far as possible, and the use of PostScript runs counter to this goal, we would like small or at least compact file formats. If we cannot keep our materials small, we should at least keep them tidy.

## **Marketing**

The marketing facet of Internet publishing is clearly the one that is most developed. Moreover, our ability inexpensively to contact a large number of potential customers provides a quantum change in the costs of publishing. Indeed, we could not consider self-publishing the Slessor diaries for nearly 25 years because of the costs of conventional advertising and mail-outs. Electronic mail (email) is cheap, even if it does not reach the entire universe of potential customers, and even if many ostensible users of email do not in fact check their "in" box on a regular basis. A variety of indicators suggest, however, that email is becoming extremely common across a wide segment of society and the "mail" can take on a variety of forms such as images, sound bits and movie clips (Carrier and Georganas, 1995).

A more difficult issue is that we must target our use of email very carefully. Blanket email transmission ("spamming") has been vociferously and justifiably criticized. Our own attitude to unsolicited advertising email is that it is acceptable and even welcome if the messages are polite, very short, and not too frequent. The recipient must also be a likely customer and not just a random email address. This implies that the marketer must have a list of email addresses of potential customers. To avoid large numbers of "returned" messages, this list must be maintained. Our experience is that such exercises require a lot of work.

An alternative is to post messages on news groups or to send short publicity items to electronic newsletters. We found that this was highly effective in attracting attention to SCPC. Moderated newsletters and news groups are more likely to be read than automatic list servers.

We believe that announcements of new publications should be in the form of very simple statements, for example:

### **NEW ELECTRONIC BOOK**

The Diaries of Arthur Slessor -- Sierra Leone 1888-91

Nash Information Services Inc. announces that these diaries are now part of our series First Person History. For information and a sample, link to <http://www.synapse.net/~nis/fph.htm> or send email to [mnash@synapse.net](mailto:mnash@synapse.net).

We strongly recommend that further information be available on World Wide Web (WWW) pages. Since, however, not all Internet users can browse WWW pages, we also want to have prepared texts of information suitable for rapid emailing. Such incoming email provides us with addresses for future announcements.

In our experience, WWW pages designers have been far too enamoured of fancy graphics. These take a long time

to load and display, and many sophisticated users of the Internet no longer display them automatically. The Lynx browser is text-only. Thus efforts to create highly detailed graphics may be wasted. Worse, if part of the "message" is included in a graphic, it will never reach its intended audience. Colleagues who have been marketing software on the Internet have stated that users will not wait to read pages that take longer than 10 seconds to load.

Our approach to graphics, illustrated in <http://www.synapse.net/~nis/whoarewe.htm>, is to include graphics that have been modified so that they are still visually attractive but require limited storage space and transmission time. Our photographs were scanned, then modified to reduce them from over 100K bytes in colour to 11K bytes in grey-scale. This does, of course, require effort. We spent a couple of hours "playing" with the image parameters in a graphics program.

We would like "Web-surfers" to encounter our advertising pages in ways other than by our direct suggestion. There are at least two tactics here:

Our main or Home Page needs to be indexed in the databases used by various "search engines", that is, WWW sites that allow users to seek material that may interest them. We have found that our pages in the directory <http://www.synapse.net/~nis/> are in several of the search databases (Lycos, Infoseek, OpenText). However, <http://macnash.admin.uottawa.ca>, our experimental server that is older than our Synapse pages, is more difficult to find because we did not "register" it with any search engines.

Our site(s) should include some useful or interesting information, software or graphics to attract browsers and to bring them back occasionally (Tetzeli, 1996). The prevailing opinion is that the appearance and content of the pages should change frequently. We agree that the content needs to change, but changing the structure and appearance too frequently may make it difficult for human browsers to find their way around our pages. We have been running the *DATAMOVE* project that attempts to provide pointers on how to translate various physical and logical file formats (Nash, 1993). There are also technical papers on the MacNash site.

Our WWW site should provide ordering information and possibly an order form (see next section). We are convinced that for any of the materials we are intending to publish in this way that there should be a sample or "teaser" for the product. This should offer a clear indication of what the customer is going to get in the full product without giving it away. The samples also attract visitors to the WWW site.

It is important that our information be available all the time. Thus, it is useful to have multiple sites as a hedge against machine or network breakdowns, but we must admit that this increases our maintenance workload. If our site(s) become well-known within particular groupings of Internet users, we have more chances of reaching likely customers. Martin (1996) calls this "seeking communities".

### **Order reception and fulfillment**

At the time of writing (March 1996), complete Internet solutions that allow automation of the order fulfillment and payment processes are not yet in place. There are many experiments in progress. Major credit card organizations (VISA and Mastercard) are apparently planning to align their protocols to standardize the handling of credit card payments digitally.

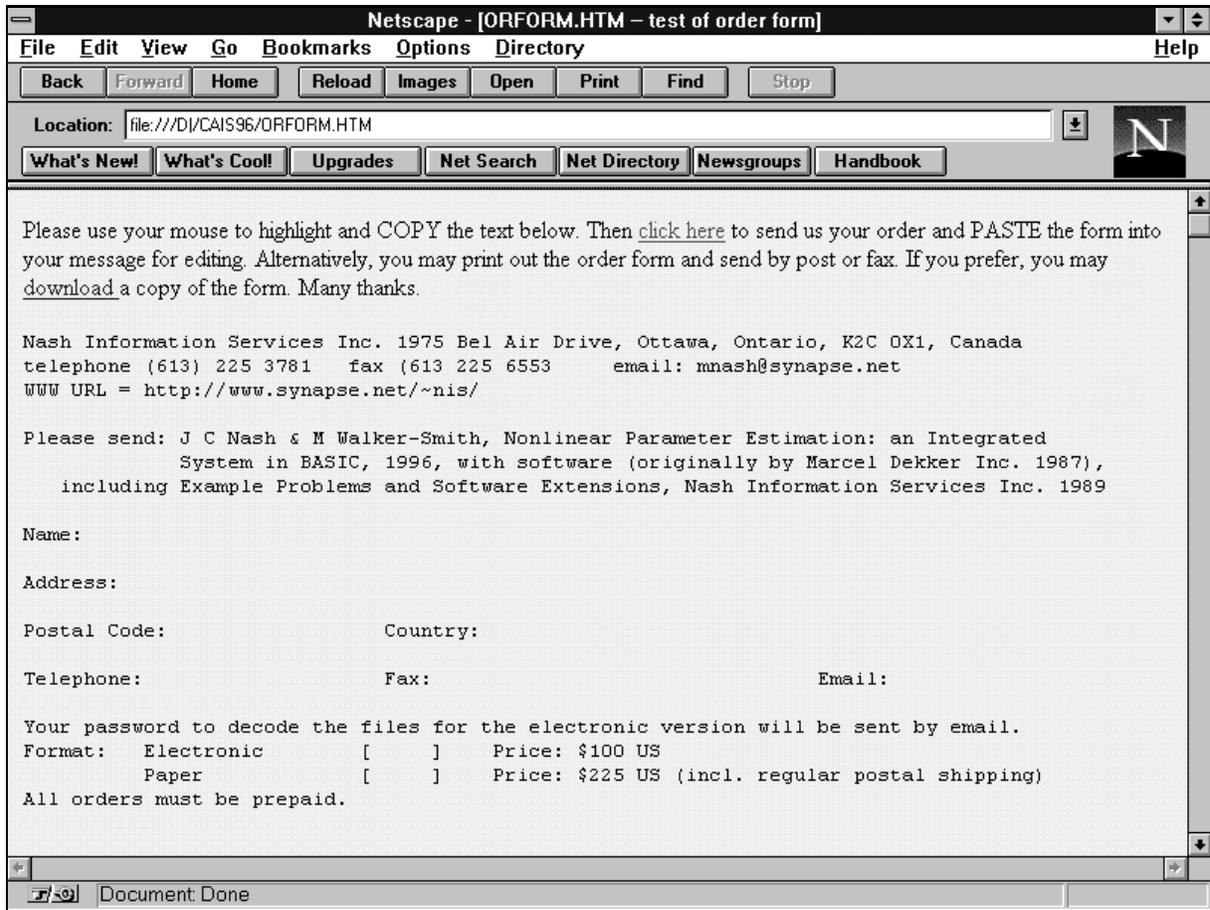
Many WWW sites already have order forms for merchandise that is then delivered by conventional means. The mechanisms to handle such order forms vary, but the principal approach appears to be a form expressed in the HyperText Markup Language (HTML). When the user submits his or her information by a mouse click on the "submit" button, information is transmitted to the WWW server and processed by a special program. The protocols for this process are collectively referred to as CGI (Common Gateway Interface) scripts. These are programs that are typically platform dependent and require technical expertise to prepare. By contrast, it is relatively easy for a computer-literate user to prepare HTML pages.

However, CGI scripts do allow publishers to conduct transactions with customers. Until Internet credit card handling becomes common, each publisher will have to provide a way to record numbers and process them. Unfortunately, until various issues relating to encryption technology are resolved, customers are likely to be reluctant to transmit their card numbers, even if this is no more risky than giving the same numbers by voice or facsimile.

While CGI scripts offer the most flexibility in handling orders and payments in ways that are convenient to customers, this comes at a high price. We need to program and test quite sophisticated programs. Generally we will want to have a dedicated and private Internet node, or else will have to pay a premium rate for service. Our provider currently charges \$30 per month for regular access, including 5 MB of space for files. Access to CGI capabilities, although it adds 45MB of storage space, also adds \$50 per month in costs. This reflects the added difficulty and potential for trouble that such facilities entail. On the experimental server MacNash, the MacHTTP server software mentions CGI scripts in the release notes, but includes no documentation nor examples of how to use them.

A much simpler mechanism is to provide "mailto" links in our WWW pages. This offers the user a blank message screen that can be filled in and sent. (Some users do not set up their browser so this will work properly, which can be a hindrance to our marketing efforts.) Ideally, of course, we want different messages to be sent depending on the product or information sought. We are experimenting with the use of FTP links that download an order form for each product that the user can then "quote" in a message sent by a "mailto" link. We can also use a simple text order form that the user can cut and paste into his message. While we can envisage downloading sophisticated programs, with appropriate versions for different platforms, we presently propose using only simple text files. Figure 1 provides an example. Note that it identifies and gives contact information for the publisher.

This allows the customer to use traditional means for ordering, including toll-free telephone numbers if the items for sale are of sufficient value. While it does not automate the order fulfillment process, it is a beginning. Figure 1. Example order form Web page that is an alternative to CGI scripts.



For the types of products we wish to publish, the customer has a choice of receiving purchased items either as machine readable documents that are ready for display or printing as printed material with a simple binding, shipped via regular post or courier. Clearly, we should price our products so that we fully recover all costs associated with the extra handling and materials that the paper forms of our products require.

In publishing *SCPC*, we noted that there were many more copies downloaded than were paid for. Though we asked for a \$10 license fee for a complete book, we were asked for "discounts" for a variety of reasons. (We actually offered a 50% discount to one student who requested it, then did not even bother to send in the reduced amount.) In part, we believe the psychology of publishing specialty items is such that we should have very much higher list prices. We are, of course, still at liberty to offer substantial discounts for whatever situations we deem appropriate.

Naturally we cannot continue the "shareware" model for our products. It appears in general that the rate of compliance for payment is 1% or less (Finn, R. 1995). (We actually did slightly better than this, 10 out of 750.) For the future, we intend to ensure payment is received before the materials are "shipped", that is, sent by post or made available for FTP downloading. Ideally, we want this to be very easily organized and ultimately want it to be automatic. The obstacles to automation are the development of programs to organize the customization of the files the lack of scriptable mechanisms to upload files to our servers, that is, programs that will automate the FTP transfer from our private machines to the servers. Such facilities are, we know, feasible. They are simply not common, nor part of generally available software. Access to the same facilities that permit CGI scripts to be run would open such possibilities.

Thus the obstacles to reducing the human element in preparing materials for download are solvable given enough expenditure of money and effort, but the total costs are difficult to estimate since they depend on the particular configurations of the machines to be used. Furthermore, any programs and procedures developed may not be directly transferable to other Internet providers. A more satisfactory solution requires that we have our own Internet node, which is still a relatively expensive proposition while our list of publications is small.

On our experimental server, we can set up private directories with password control. We have used this facility together with password-access FTP to enable exchange of files with colleagues. This is one possible mechanism to limit access to the files we wish to distribute. [University regulations forbid the use of the MacNash machine for commercial transactions. We can and do, however, use it as a testbed for ideas and programs.] We cannot do this easily on the Synapse server because the overheads of authentication of connections are high enough that the operators have decided not to activate these features of the server software. [We could, of course, program our own within the CGI capabilities.]

A simple and quite inexpensive option is to encrypt the files to be downloaded and then send the key to unlock the files to paying customers. Such encryption would not be expected to withstand serious challenge, and sending the key by email would be open to losses. However, as an obstacle to theft it would likely be satisfactory for our needs, since serious cheaters could simply pay for one copy then distribute it widely as with pirate editions of CDs. It is a threat that is mainly countered by the anticipated dispersed market for the products. Under this assumption, we could give all customers in some period the same password. That is, we would only have one or two versions of our file for download and rely on the geographic separation of likely customers to inhibit unauthorized downloads.

One advantage of an encrypted file is that we can provide checks so that the customer can be sure they have successfully downloaded files before they pay. Of course, they cannot use the material until they get the key to decrypt it. Verification would probably rely on file sizes and cyclic redundancy checks (McNamara, 1982).

We plan to offer a printed form of our publications, though this is **not** the primary format. We believe that there will be people willing to pay a premium to receive documents on paper. To keep costs low, we will need to automate selected parts of the printing and shipping process to minimize human effort.

## **Future directions**

We expect it to still be some months or years before the ordinary user of the Internet both trusts and finds commercial transaction facilities convenient. Furthermore, the benefits may only be realizable if a sizeable investment can be made. A large investment is not sufficient, however, to guarantee success, since customers will shun poorly implemented sites and simply turn away from any awkward operations.

Despite the obstacles, we intend to persevere with the development of content suitable for Internet publishing. Content takes much longer to prepare than most "surfers" comprehend, even if the basic source documents are in hand. The *First Person History* endeavour under which the Slessor diaries will be released should have several titles within the next few months. *SAA*, the Scientific and Statistical Software Source and Advisor, has been put on a back burner but is nonetheless still alive and is the umbrella for re-publication of both *SCPC* and Nash and Walker-Smith (1987, 1989) along with, we hope, a number of other works. The *DATAMOVE* initiative, while primarily a vehicle for organizing tools for our own purposes, serves as a point of attraction on our WWW sites and could also be a framework for the publication of specialty software for file translation. All of these initiatives are open to collaborations with others. Indeed, one of the great opportunities offered by Internet is the publishing of a vast panoply of works having individually a tiny market, though together they represent a large volume of business. We welcome comments and collaborations.

## Acknowledgement

The research of J C Nash has been assisted by NSERC and SSHRC grants.

## References

Carrier, S. and N. D. Georganas. 1995. "Practical multimedia electronic mail on X.400." *IEEE Multimedia*, vol. 2, no. 4, Winter, 12-23.

Finn, R. 1995. "Access to freeware and shareware archives." *Scientist*, vol. 9, no. 15, July 24.

Martin, Michael H. (1996). "Why the Web is still a no-shop zone." *Fortune*, vol. 133, no. 2, Feb. 5, 127-8.

McNamara, John E. (1982) *Technical Aspects of Data Transmission*, Bedford MA: Digital Press.

Nash J. C. and M. M. Nash. 1995. "Internet book publishing -- a case study." in *Connectedness: information, systems, people* (ed. Hope A. Olson & Dennis B. Ward), Canadian Association for Information Science, Proceedings of the 23rd Annual Conference, School of Library and Information Studies, University of Alberta, 108-116.

Nash J. C. and K. Price. 1979. "Fitting two straight lines." *Proceedings of the Tenth Interface: Computer Science and Statistics*, American Statistical Association, Washington, 363-367.

Nash J. C. and M. Walker-Smith. 1987. *Nonlinear parameter estimation: an integrated system in BASIC*. New York: Marcel Dekker Inc.

Nash J. C. and M. Walker-Smith. 1989. *Nonlinear parameter estimation: examples and software extensions*. Ottawa: Nash Information Services Inc.

Nash J. C. 1993. *Moving and manipulating computerized information: strategies, tactics and tools*. Ottawa: University of Ottawa, Faculty of Administration Working Paper 93-39.

Okerson, A. S. and J. J. O'Donnell. 1995. *Scholarly journals at the crossroads - a subversive proposal for electronic publishing*. Washington, DC: Association for Research Libraries.

Peters, J. 1995. The Hundred Years War started today: an exploration of electronic peer review. *Internet research: electronic networking applications and policy*. vol. 5, no. 4, 3-9.

Suwalski, W. 1988. *WED -- a text editor for MS DOS computers*. Nepean, Ont.: Labcomp Computing.

Tetzeli, R. 1996. Getting your company's Internet strategy right. *Fortune*, vol. 133, no. 5, Mar. 18. 72, 74, 78.