The Benefits of Electronic (Paperless) Publishing

1. Transmit documents to your next-door neighbor around the world in minutes. Your audience is anyone with access to a computer.

2. Update documents as often as you like. There's no need to republish all your information, making many versions of the same document.

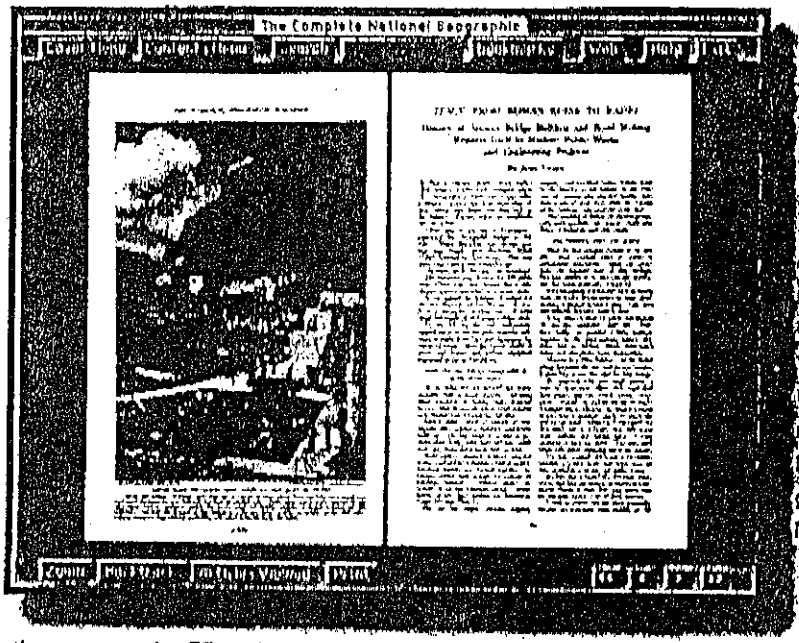
3. Save some trees. A vast amount of information can be stored on a single disk rather than stacks of paper. Recycle the disks by reusing them. Print what you need from your own printer and reduce waste.

4. Publish yourself. Everyone can become a published writer. You don't need anyone else to produce your work. Get your ideas across the globe while sitting at your desk.

white with some halo around it. This is especially true in the older documents, scanned from yellowed paper. We did as much as we could to correct this. The purpose was to create a consumer product which forced us to make compromises in quality."

National Geographic decided against full-text search. "We thought it was overkill for the average user," says Stanton. "We had an index in the library where they looked at every page and assigned five key terms to every page. They did this by hand. Blessed with this in-house index, we thought it would more than suffice.

"We considered OCRing all the text but we needed 100% accuracy. Having 99% wouldn't be up to National Geographic standards. National Geographic wanted to preserve the look, feel and style of the magazine. Having the page image did the best job."



National Geographic published the magazine's first 108 years on a CD-ROM entitled, *The National Geographic Society's First 108 Years.* Every page from some 1,235 issues was scanned. They used Ledge Multimedia, a division of Dataware Technologies (Cambridge, MA 617-621-0820).

Ledge used Dataware's proprietary ADL database. The database is invisible on CD format. Their index search database was used to create a data retrieval library which sits on every CD.

Each CD also contains its own navigation database of all the images, giving related pages and articles. They are all in directory structures based on month and year. Each image is identified by type and by page number. The database tells the user what CD to insert in a search.

The images that were sent on CD from DAD were scanned for millions-of-colors display. Because this is a consumer product National Geographic wanted the images to also be made viewable for a 256 color display. To do this a second version of the images had to be made with a specific color palette for each set of images (each two page spread). This second palette is invisible to the user. If their monitor has a 256 color display the user doesn't know the difference. This display feature was written specifically by Ledge for National Geographic.

Ledge wrote multiple versions of the project for approval by National Geographic. They recorded the CDs using Adaptec Toast Software and a variety of internal CD-R drives.

"The biggest challenge was dealing with the magnitude of data, says Kryger. "Once everything was indexed we had to look through every page to verify the indexing and navigation was correct."

Their second major task was to index the advertisements that appeared in the magazines. This was done chronologically. After 1970 National Geographic included regional advertisements. This made it difficult to index. Ledge wrote a Visual Basic application for the indexing. They had to go through the scans on the CDs to locate the odd regional adds and replace them.

Users can search across the CD-ROMs through every issue by criteria such as title, subject, keyword, place, name, contributor and date. Ads can also be searched for separately by subject and date.

"We were very fortunate to find Ledge," says Stanton. "They were able to integrate the scanning with the index table from the magazines. They also integrated our own index material consisting of tens-of-millions of individual words or phrases derived from our Library Sciences Group who carefully indexed all of these pages for years. It was a major boost to the project that this in-depth index for the text already existed. We were also able to have a hyper index for related subject or contributor."

"The project was reasonably cost effective," says Stanton. "Our biggest trade off was the compression which had to be tight. We used JPEG which worked the best on images and not

National Geographic also considered publishing on the Web. They looked at what it would take to download an entire article which would typically be about 20 pages long. That would be half to three-quarters of a megabyte to look at one article. "We thought CD would be a perfect medium," says Stanton. "The only downside is the 30 CDs. We're looking into DVD to reduce this.

"We are very pleased with Ledge. They will do other projects for us in the future. We were relieved to find a team of solid engineers with a great approach. They are a sophisticated group which had experience with commercial CD-ROM publication. They really cared about the work they were doing and they fell they in love with the project without any seduction from us." ■

The publishing can be done easily in house with a robust scanner, a CD-Recorder, or the right size Web server for Internet publishing. The conversion work can also be sent to service bureaus and electronic publishing houses. Many service bureaus offer Adobe PDF file conversion and electronic publishing services. PDF plays a big part in the electronic publishing arena. The PDF format makes documents viewable on any platform while keeping all of the original document elements including layout, fonts, imbedded video and audio.

Another purpose to electronic publishing, besides distribution, is being able to preserve old and fragile paper documents safely. This gives access to the public without worrying about the wear and tear to irreplaceable documents.

National Geographic Publishes 108 Years On CD-ROM

National Geographic completed a project last fall to publish the magazine's first 108 years on CD-ROM entitled, *The National Geographic Society's First 108 Years*. The 30-CD collection spans over 10 decades. It includes the magazine's first issue in October of 1888 and continues up to the December issue of 1996. National Geographic published their issues to CD-ROM with the goal of preserving the information and making it available to the public.

Every page from some 1,235 issues was scanned — more than 190,000 printed pages in total. This included text, graphics, photography and advertisements. National Geographic started the testing for the project and coming up with requirements in February 1996. The work began in September and the CDs began shipping in August 1997.

"The approach was to scan the page and treat it as electronic microfiche. We looked to preserve the actual page image and all the photos," says Tom Stanton, director of CD-ROMs for National Geographic Interactive.

National Geographic created the CD collection for consumers. The CDs had to be able to run on a 486 66 PC with 8 MB, running Windows 3.1. "We created the product without high requirements realizing that not all users can afford to upgrade to Windows 95 or the latest thing," says Stanton.

"We went out locally to find a scanning service bureau in DC. Most companies were doing work for government applications, primarily a black and white science. We wanted to scan 24-bit color at a minimum with high photo quality. Creating a consumer product, tion because of the limits of display." National Geographic came across Ledge Multimedia, a division of **Dataware Technologies** (Cambridge, MA 617-621-0820).

Ledge Multimedia also worked with **Document Automation Development** (Overland Park, KS 913-663-4323), a provider of document automation services including electronic publishing and scanning. Document Automation Development (DAD) had the monumental task of scanning all 1,235 original issues. From there the scanned images would go to Ledge Multimedia for merging the scanned material and multiple indexes. Each image is a two page spread from the magazine.

The two companies developed a unified standard for use throughout the project. Test pages were scanned at various settings until they found the right settings to optimize the magazine's pages including photographs and images.

DAD used a collection of five software applications created in-house, collectively called DocuTrak. DocuTrak has a proprietary process for workflow management that allowed DAD to accurately index and track some 300 total GB of images and information. Three complete sets of the magazine were delivered on three forklifts direct from National Geographic. The first set was the master set. The second was a safety set, in case any pages in the master set were unusable with folds or stains. The third set was a back-up in case of a total catastrophe in the process.

"We used DocuTrak to index every page of every issue from front cover to back cover, including foldouts," says Vince Pingel, president of DAD. "An operator turned the pages of each issue and entered the information into the index format by hand. Once every page was indexed, the information went into a database."

The binding was cut off the master set and DAD began scanning the issues. DocuTrak gives on-screen prompts to the scanner operator that tell the person exactly what pages they should be scanning. DAD reviewed every scanned image as part of their control on the scanning process.

DAD's use of DocuTrak in their work with National Geographic let them eliminate a lot of the manual work and minimize the risk of human error. One of DocuTrak's five components is an indexing module that drives the scanner and assigns the images file names. Another is the workflow component that drives the entire process. A third is a quality control piece that does a page by page check of every image burned to CD-R against

the original index. A report generator itemizes the entire product and database in a spreadsheet fashion. This gets passed on to Dataware for them to build on.

To automate the process of giving the images file names, DAD had to create a structural index, assigning page numbers to all the pages that were numbered and unnumbered, including foldouts and ads. "We took a page-by-page account from all the issues to figure out how to fit this into a file naming scheme," says Pingel. "This is where DocuTrak came in. We created our structural index of the magazine and put that into DocuTrak. DocuTrak drove the scanning process and guided the operator."

DocuTrak gave prompts to the scanner operator, automatically named the file according to the structural index and placed the file in the proper directory. DAD automated the entire scanning process down to one person. They used one operator running two scanners at a time. The operator was freed up to do this because Docutrak automatically set all the scan settings.

File compression, image rotation, moving the files across to the network to the burner and staging of the CD-R was all done automatically by DocuTrak.

> Another purpose to electronic publishing, besides distribution, is being able to preserve old and fragile paper documents safely.

DAD created about three GB of scanned images a day. These were compressed using Pegasus Software and staged for CD mastering. Two complete sets were created, one for Ledge Multimedia and one for DAD's backup. There were 644 CDs per set — 1,288 CDs in total were authored in house. One CD holds two issues in uncompressed BMP format and JPEG compressed format. To record the issues onto CDs, DAD used a Panasonic CD-recorder, an HP CD-recorder, Sony CD-R media and Adaptec CD creation software.

After mastering, each CD was checked again for errors by physically comparing each scanned image with the original page from the safety set of magazines. "We did everything we could to make sure every page was accounted for and the scanned pages were error free," says Pingel.

The complete set of images on CD was sent to Ledge Multimedia to be merged with the vast index created by the Library Science Group at National Geographic and Ledge's index search database and image navigation database format.

"We worked closely with National Geographic on the design of the project," says Lisa Kryger, executive producer from Ledge Multimedia. "We helped them come to terms with what they could and could not do. They wanted all the documents kept in the same format they originally appeared in with the same index used in the magazine."

The Benefits of Electronic (Paperless) Publishing

1. Transmit documents to your next door neighbor around the world in minutes. Your audience is anyone with access to a computer.
2. Update documents as often as you like. There's no need to republish all your information, making many versions of the same document.
3. Save some trees. A vast amount of information can be stored on a single disk rather than stacks of paper. Recycle the disks by reusing them. Print what you need from your own printer and reduce waste.
4. Publish yourself. Everyone can become a published writer. You don't need anyone else to produce your work. Get your ideas across the globe while sitting at your desk.

Ledge used Dataware's proprietary ADL database. The database is invisible on CD format. Their index search database was used to create a data retrieval library which sits on every CD.

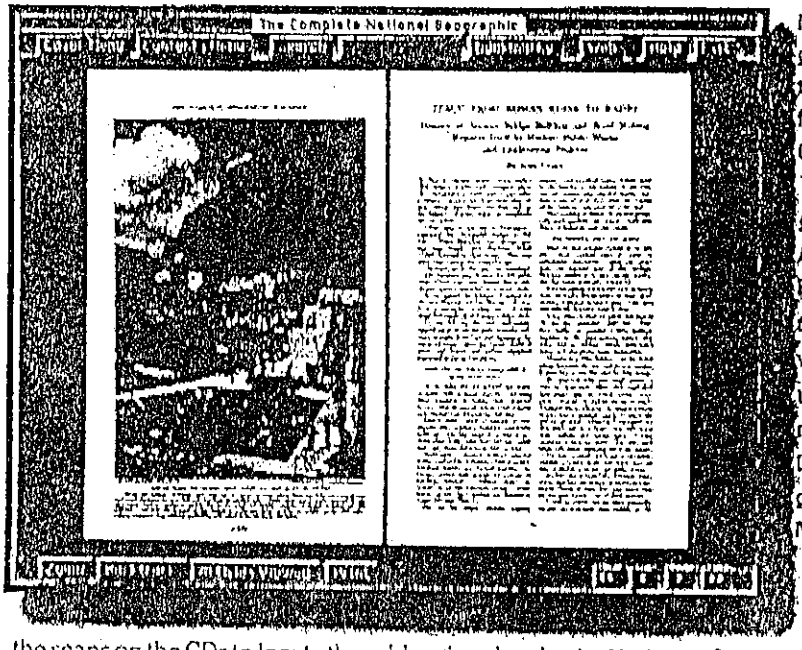
Each CD also contains its own navigation database of all the images, giving related pages and articles. They are all in directory structures based on month and year. Each image is identified by type and by page number. The database tells the user what CD to insert in a search.

The images that were sent on CD from DAD were scanned for millions-of-colors display. Because this is a consumer product National Geographic wanted the images to also be made viewable for a 256 color display. To do this a second version of the images had to be made with a specific color palette for each set of images (each two page spread). This second palette is invisible to the user. If their monitor has a 256 color display the user doesn't know the difference. This display feature was written specifically by Ledge for National Geographic.

Ledge wrote multiple versions of the project for approval by National Geographic. They recorded the CDs using Adaptec Toast Software and a variety of internal CD-R drives.

"The biggest challenge was dealing with the magnitude of data, says Kryger. "Once everything was indexed we had to look through every page to verify the indexing and navigation was correct."

Their second major task was to index the advertisements that appeared in the magazines. This was done chronologically. After 1970 National Geographic included regional advertisements. This made it difficult to index. Ledge wrote a Visual Basic application for the indexing. They had to go through the scans on the CDs to locate the odd regional adds and replace them.



National Geographic published the magazine's first 108 years on a CD-ROM entitled, *The National Geographic Society's First 108 Years.* Every page from some 1,235 issues was scanned. They used Ledge Multimedia, a division of Dataware Technologies (Cambridge, MA 617-621-0820).

Users can search across the CD-ROMs through every issue by criteria such as title, subject, keyword, place, name, contributor and date. Ads can also be searched for separately by subject and date.

"We were very fortunate to find Ledge," says Stanton. "They were able to integrate the scanning with the index table from the magazines. They also integrated our own index material consisting of tens-of-millions of individual words or phrases derived from our Library Sciences Group who carefully indexed all of these pages for years. It was a major boost to the project that this in-depth index for the text already existed. We were also able to have a hyper index for related subject or contributor."

"The project was reasonably cost effective," says Stanton. "Our biggest trade off was the compression which had to be light. We used JPEG which worked the best on images and not white with some halo around it. This is especially true in the older documents, scanned from yellowed paper. We did as much as we could to correct this. The purpose was to create a consumer product which forced us to make compromises in quality."

National Geographic decided against full-text search. "We thought it was overkill for the average user," says Stanton. "We had an index in the library where they looked at every page and assigned five key terms to every page. They did this by hand. Blessed with this in-house index, we thought it would more than suffice.

"We considered OCRing all the text but we needed 100% accuracy. Having 99% wouldn't be up to National Geographic standards. National Geographic wanted to preserve the look, feel and style of the magazine. Having the page image did the best job."

National Geographic also considered publishing on the Web. They looked at what it would take to download an entire article which would typically be about 20 pages long. That would be half to three-quarters of a megabyte to look at one article. "We thought CD would be a perfect medium," says Stanton. "The only downside is the 30 CDs. We're looking into DVD to reduce this.

"We are very pleased with Ledge. They will do other projects for us in the future. We were relieved to find a team of solid engineers with a great approach. They are a sophisticated group which had experience with commercial CD-ROM publication. They really cared about the work they were doing and they fell they in love with the project without any seduction from us." ■

August 17, 2001

Dear Jerry and Idaz,

I have the one page synopsis of Norm's points in hie brief about CD108. I do have a few comments:

1. Where he discusses "minimal software" needed, I'd like to be sure that it's understood that quite a few things have to be in place before the CD can be used on either a PC or a Mac. Those things have to be both on the consumer's computer and also on the CD itself. For instance, for the consumer's computer to activate the CD it's necessary to have an operating system, several graphics files, specialized files to access a CD itself, and a CD drive either inside or outside the computer. A lot has to happen in order to successfully use the CD108 product. It's not minimal at all. It's substantial.

2. Bullet "Word-search capability:" CD108 has much more than a word-search capability. Some of it is covered in the next Bullet about advanced searches. But it's important to point out that this is a very sophisticated set of features that NGS has included in its menu. You can search on names, article titles, topics, key words, and a host of other things to get to the printed article you want to see. You can do combination searches and type in all kinds of queries to make a collection of topics, titles, names, etc. I believe this might be spelled out more clearly with examples so the court actually gets a feeling for just how sophisticated the NGS search facility really is. There is no way to do any of these things from the printed magazines.

3. I am not sure what the last bullet means... The one about downloading something into the computer's hard drive. I just don't know what can be downloaded and saved using the NGS search engine. I know one can easily go into the CD without using the search engine and get at all the pictures and texts and download. But this is different from using what comes up on the screen.

4. Perhaps it would be good to include a bullet about being able to access each individual picture on the CD and download it into your own computer and use it anyway you like.

5. The bottom typed line is about the 1997 copyright notice. You have not put in the 1997 in the right place. The actual line that is printed at the bottom of every page a consumer prints is: © National Geographic Society 1997. Be sure when using this line that you get in the 1997 date and put it in the right place.

That's it. Regards, Fred Ward