

Warnock on PDF: Its Past, Present and Future

BY BERND ZIPPER

Ten years after 'Carousel' was demoed to a Seybold Seminars audience, PDF has become the standard file format for print-oriented documents. But its roots go back to 1984. Adobe founder John Warnock tells how it all came to be.

Bernd Zipper: John, please tell us the origin of the idea that developed into PDF.

John Warnock: It's probably going to be surprising to you, but the first real germ of the idea was in 1984. We were about to bring out the Apple LaserWriter and we wanted a bunch of samples to show. So I hand-programmed an IRS tax form—I actually brought it along to show anyone who was interested—but I hand-programmed this in PostScript. It's a difficult document to hand-program, so you make a lot of subroutines and utilities to help you image the page. We ran it on the original LaserWriter; it took 2 minutes and 45 seconds to print. And I remember Steve Jobs said: "No, we can't have any page that takes that long."

So I started thinking. There was a property of PostScript that most people didn't know about—well, people who really knew PostScript knew about it—but you can redefine the operators to have different semantics than the original operators. So I took all of the basic graphics commands in my original program, reprogrammed just a capture of their parameters and wrote those parameters out to a file. To use a programming term, this "flattens" the file out. The resulting file printed on the LaserWriter in 22 seconds. So we used to affectionately call this the "graph binder," and it stood for binding the graphics operators in PostScript.

That idea sort of went to sleep and, in 1991, it started to become clear that PostScript was going to become a very big standard in the printing industry. So I thought about it and I started also thinking about the proliferation of networks. This was before the explosion of the Internet. We were thinking about local area networks and we were thinking about wide area networks; we weren't yet thinking about the community of the Internet. I said, "What the world really needs is a way to send documents across platforms." I thought that, using the original graph-binder technique, we could take any PostScript file and flatten it, and so make it easy to interpret and easy to render. It didn't have to have the full PostScript language. At that time, Display PostScript wasn't gaining traction and it looked like

it wasn't going to gain traction. So I wrote a small paper—I'll give you a copy—called the Camelot paper.

Zipper: Thank you, John. And this paper is the first definition of PDF?

Warnock: This is the very first paper that discusses Acrobat. And it essentially lays out the strategy and the market opportunities for PDF.

We then got a very small tiger team together. That tiger team consisted of about three programmers and a couple of the senior computer scientists. The three programmers were going to build a prototype of a rendering engine that could render on the screen these files that came out of this graph-binder process. They made it quite fast—remarkably fast for the machines that were available at that time. And that gave us encouragement that we could really do this process.

Then I went to a fellow named Doug Brotz who had been heavily involved in the PostScript inter-

Germ of an idea. While hand-coding PostScript to make this tax form print faster on the original LaserWriter, John Warnock developed the key concept behind PDF.

preter since the very first days. I said, "What I need is a version of the graph binder that is inside of a PostScript interpreter and is real robust." The concept was that it could deal with all of the aspects of Type 1, Type 2 and Type 3 fonts and be able to output these various files. He started to work on that problem.

The other part of the work was done by Peter Hibbard. He was one of the most senior computer scientists within Adobe and he was very familiar with document structures. I asked Peter to build a flexible file format that was extensible and that could take advantage of the flattened files. Flattened files have advantages over PostScript files; because the pages are independent, you can build pointers to specific pages, you don't have all of the conditional operators and you have a lot of static information about the document. Peter designed a thing called COS that was the foundation for the PDF language and PDF data structures.

Zipper: So we can say that PDF is ten years old this year [2001].

Warnock: That's right.

Zipper: That's great, because Stephan Jaeggi, Jim King and I talked about this and we were not sure.

We said the man who got the idea and developed it has to settle the question.

Did the people at Adobe really understand the logic behind PDF, when you first introduced it?

Warnock: Well, I did. So did people on the PostScript team—Doug Brotz, Peter Hibbard and a few more people. But getting the marketing organizations and the sales organizations to understand it was very slow going. Well, we didn't have the Internet then.

Zipper: You introduced Acrobat, initially called "Carousel" at Seybold Seminars in 1991. What was the first reaction of the audience?

Warnock: The people who were in publishing and who were really involved in the process knew the implications of having a format where you can send a document across platforms and across network structures and preserve all of the fonts, because it's not a new problem. It's a problem that's been around for 30 years. SGML, XML and so forth were all attempts to create abstract documents so that they could be communicated. What occurred to me was that there is an opportunity and a time and a place where all of the print drivers could go through a PostScript print driver and you could essentially capture all of the documents without getting the application programs' permission. In other words, you didn't have to go get cooperation from Microsoft, you didn't have to get cooperation from Quark, you didn't have to get cooperation from anybody! You could just capture the PostScript files and then go back later and put in structure.

Zipper: Adobe's marketing decided at that point to initiate a program around a theme like "the paperless office."

Warnock: Yes.

Zipper: Did you believe in the paperless office at that time?

Warnock: Well, no, because people still like to read books; they still like to read paper. But what I did believe in was communication of documents without paper; that is, getting the document from point A to point B without having destroy a bunch of trees.

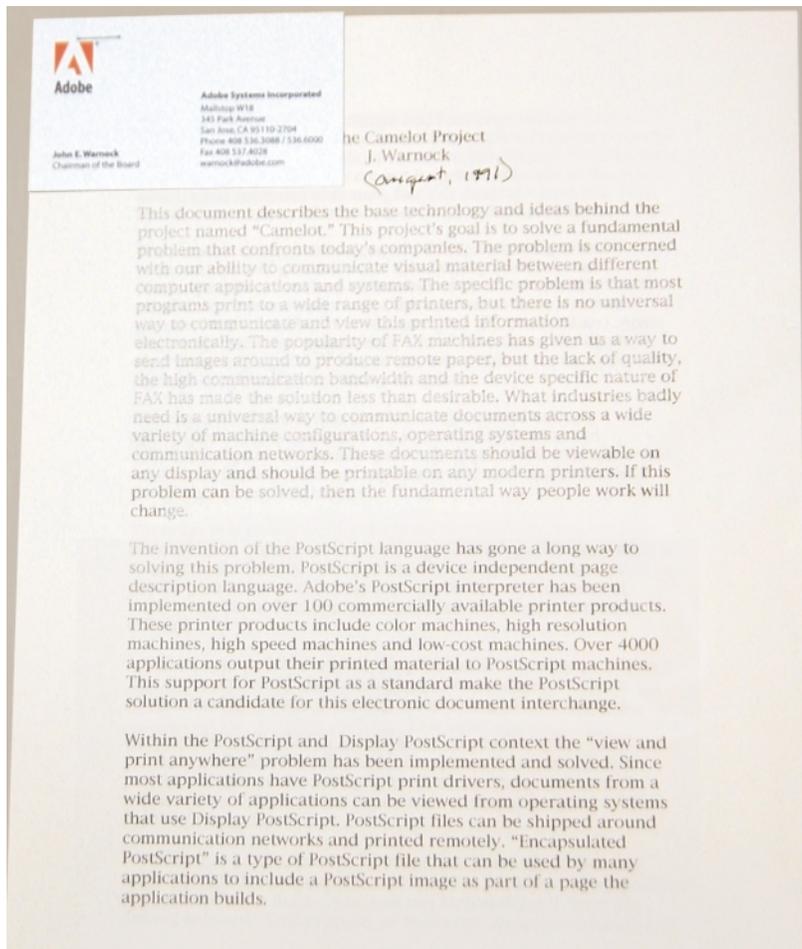
Zipper: And it wasn't very easy. It took you perhaps five years to make Acrobat and PDF into a real business.

Warnock: Yes.

Zipper: That's not normal, nowadays, for business and software development.

Warnock: Do you think it's long or do you think it's short?

The "Camelot" paper. In 1991, Warnock first described the market opportunity for Acrobat in this document.



Zipper: I think it's long.

Warnock: Yes.

Zipper: If you compare that with other products—for example, if you compare that with GoLive—these products have to bring in money very fast and there is no chance to build up a perfect foundation to let it grow.

Warnock: Quite frankly, we thought people were going to get it much faster than they did. I thought they were going to understand the applications of Acrobat much more quickly than they did, but they didn't, and it started to grow very, very slowly. The growth was always there; it was always on an up-slope, it never started to go down. But it was very slow. There was a lot of pressure in the company to kill the product because it wasn't an economic success, but I was absolutely convinced that this was a long-term winner and did not allow that to happen.

Zipper: Did you expect that the printing and publishing industry was going to accept PDF first? Did you expect that?

Warnock: No. Actually, I thought that there were a lot of problems that had to be solved to get the printing technologies correct. I actually thought office workers and network administrators would understand it first.

Zipper: What were the major problems at the beginning?

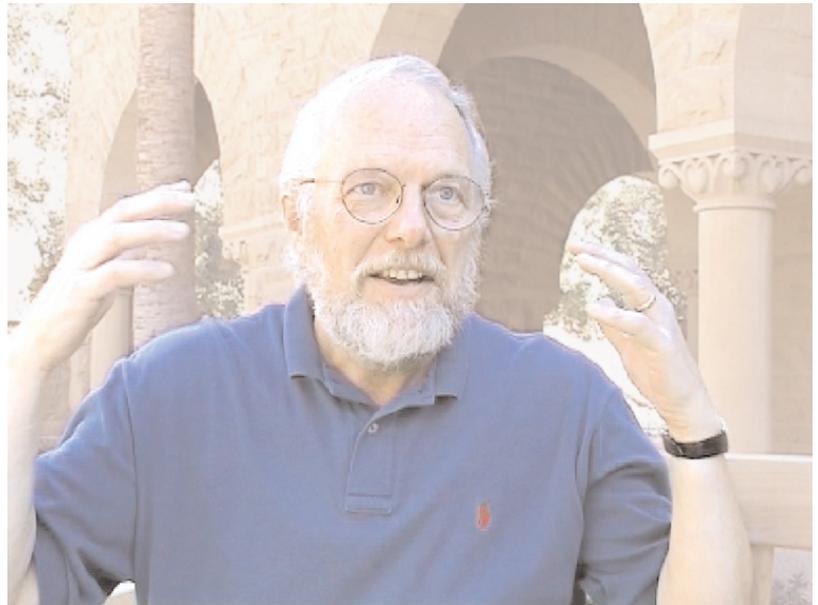
Warnock: We decided early on that there would be two products. There would be the Reader and then there would be full Acrobat. At the beginning, we charged for the Reader, and . . .

Zipper: . . . by the way, the first time I thought about it, I wondered why I should have to pay just to look to a piece of electronic paper that I can't do anything with except read?

Warnock: Well, we thought people would understand the value. It's really funny that people don't get it intuitively. It was obvious to me that this is the way you transmit information, this is the way you do slide shows and presentations, this is the way you do spreadsheets and everything else. The cost-effectiveness of it is so obvious that people should understand this. But it took a long time.

Zipper: And in developing and shaping the product, what were the major problems? You had a lot of licensed users of PostScript; I would have to believe that they would see PDF and would be a little bit jealous about that technology.

Warnock: Could be. One of the major obstacles (and, I think, one of the best inventions that was part of the early Acrobat system) was the font-substitution engine. It would mimic the widths of any font, so



John Warnock

that you wouldn't have to transmit the font along with the document. It turns out that's not used very much; people are quite happy, with Internet speeds today, to transmit the fonts along with the documents. But in the early times, that was a bottleneck, and I thought that was a particularly clever invention to figure out how to do that.

Zipper: And now, after ten years, are you still having fun seeing your product grow?

Warnock: Oh, absolutely. I'm really excited about seeing its uses broaden out to various other things, like forms. In the market place, PDF is systematically, little by little, replacing ordinary paper flow. And that ordinary paper flow is now making a transition into the electronic world—and that was the original vision. It's just that it's taken a long time and a great deal of engineering to make it happen.

Zipper: Are you still thinking about new ideas for PDF and bringing them to the company? Or do you say, OK, I did my work, let the others think?

Warnock: No, I still push pretty hard. Sometimes you'll see a little company that is doing something clever with PDF and I bring it to Adobe Ventures: "See what's happening."

It has a life of its own. It's nearly Adobe's biggest product now. I think it has a huge future if we can manage the way we add capability to it and manage the way that it's installed and the way we put value into it.

Zipper: What would you expect in the next ten years from PDF?

Warnock: In most bookstores, and at sites like Amazon, you will be able to browse PDF files before you buy. I think all serious documentation will end up

in PDF. I think all the publishers will migrate eventually towards PDF. Publishers are very slow beasts and it takes a long time, but the writing is on the wall; the Internet is a powerful communication mechanism, and it costs a lot to print a book, and it costs a lot to inventory a book. I really do believe in print on demand, and I really do believe that some sort of instant generation of books will become the rule rather than the exception.

I really do believe in print on demand, and I really do believe that some sort of instant generation of books will become the rule rather than the exception.

Zipper: There's a discussion at the moment about the technical future of PDF, and there's a close connection to XML. So I have to ask: Shouldn't PDF be totally encoded in XML? Do you see a chance for that?

Warnock: I don't see why. We have done all of the technical due diligence and, clearly, you can code PDF into XML. Does that make it easier to make an implementation? Not really. Does it add value in terms of file size? No, it probably costs file size. Could you have a single-threaded, hierarchically structured XML document that's a PDF file? Never, because the graphics rendering and the logical organization of a document sometimes are orthogonal. So you have to have a structure for the graphics and a structure for the logical organization of the document, and when you have two trees and they have to point to each other, the XML syntax doesn't help you in any way.

We've gone through that discussion and we've asked, of what benefit is it to the user for this all to be in XML? Does it make any part of the problem easier? And the answer comes back, no.

Does it add any value? Well, yes; you can say, "It's all in XML." But so what? The only benefit that may have would be a marketing benefit. Embedding XML fragments inside of PDF is not a hard problem; we do it all the time.

Zipper: What do you think about transparency in PDF?

Warnock: [Smiles.] Eventually it has to work.

Zipper: What do you think about the implementation of PDF in OS X? Is it a little bit strange, maybe? Is it a new way?

Warnock: [Smiles again.] That also has to work. Right now, there are a lot of files that OS X *can* deal with—at least, that's my understanding. I haven't

tested this myself, but I'm about to switch over to it and test it. It's really funny: When Steve started Next, he came to Adobe and we gave him this PDF/PostScript. He loved it, but when he went back to Apple, he knew he wasn't going to get this by PostScript. So he started working with PDF. It seems to me a natural trend in the way that he builds things.

Zipper: What kind of computer do you use yourself?

Warnock: I have a Titanium portable; I have a G4. I'm still a Mac person.

Zipper: So you must sometimes be very sad when you see that there are some functions in Acrobat for Windows that are not in Acrobat 5.0 for Mac.

Warnock: I'd really like the platforms to be consistent. I remember we brought out Web Capture for the PC before we brought it out for the Mac and all the Mac owners screamed and yelled, because Web Capture is really cool. So I pushed to have it included in the Mac stuff, and they finally did get around to converting it. But there was a deadline in the implementation and that's just not going to happen for the implementation that we need for Windows. The other thing is, most Acrobat users are Windows users. There's a larger group on Windows than on the Mac side, so that should naturally drive the development processes.

Zipper: PDF is independent from every application software package, and so for me, PDF is something like "data democracy," because I can take it and use it anywhere I like. Was that your intention, as well?

Warnock: Yes, absolutely. The other thing is that, when we started, there wasn't much you could do with a PDF file—just print it and look at it. It's always been my belief that over time, application programs should put more and more structure into their documents so that you can reuse the pieces. And I pushed very hard to support this in Acrobat. For instance, in Acrobat 5 you can say, "Extract the images out of the document." I was also a strong proponent of the feature that allows you to select a continuous stream of text and extract that from the document.

So I look at PDF as a repository. It is really a philosophical issue here. SGML has always had the benefit that, if you plan a bit and really think about the structure of a document, and then if you encode it in SGML, you can reuse the components almost perfectly from that document. But no one thinks ahead; they very rarely put a lot of structure into documents. The usual publishing process is to author only as much as you need and then build the document. And the challenge should be, not to increase the work at the front of the authoring process by imposing structuring on everybody, but rather to take the

myriad of documents that are around the world and figure out how to get information out of them—to build more and better tools to extract information out of them, whether it be the photographs, the graphics or the text.

Zipper: If you review all the work that you did up to now, are you satisfied with your life's work?

Warnock: I don't see how anyone in my position couldn't be satisfied. Chuck Geschke and I had no idea at the beginning that we would have a profound effect on the printing and publishing industries. We thought we'd sell a few printers here and there. Good fortune and luck have allowed us to invent and to have an environment where those inventions are really accepted and used. I'm extraordinarily proud of the accomplishments of Adobe and what the company has been able to do.

Zipper: If you got a chance to develop some more solutions, what would you want to do and what would you try to develop?

Warnock: Computers have a long way to go to be easier to use. If I think about what I have to know in order to use all the programs I use, it's an enormous amount of information that's very hard to combine. The basic trend in software is to use the power of the machine to build better interfaces to make things easier for people.

Zipper: Thank you for your time and your open answers.

Warnock: Great, Bernd. Don't forget the papers I gave to you; Camelot was written in 1991, and you should tell your readers that this was before the Internet and all these things. And you should tell them that I had the right idea ten years ago. **TSR**

Want to Learn More About PDF?

PDF has become the format of choice for many print and online applications. Yet, not all applications are well suited for PDF and working with PDF files still poses some challenges. To help publishers struggling with these and other PDF issues, Seybold Seminars is launching a new PDF Conference devoted to delving into PDF's thorniest issues. Being held February 21–22, 2002, in conjunction with Seybold Seminars New York,

the two-day conference features tracks devoted to PDF for print publishing, PDF for application developers and integrators, and PDF for government and enterprises. If you work with PDF, you won't want to miss the valuable insights, information and tips experienced PDF pros will share in each conference session. For more information on the PDF Conference and the Seybold Seminars New York program, visit www.seyboldseminars.com.

Warnock on PDF: Its Past, Present and Future	3
---	---