



INNOVATING@SUN

Innovating@Stanford University

Announcer You're listening to the Sun Microsystems Podcast Network.

Announcer Welcome to another edition of Innovating at Sun. Today's topic, Innovating at Stanford University. And now, here's Hal Stern.

Hal Stern:

Hello, and welcome to Sun Microsystems Innovating at Sun Podcast. I'm your host, Hal Stern, Vice President of Global Systems Engineering. And here, for our special 25th birthday celebration is Michael Keller, who may be the only person with a title longer than mine. Michael is the Stanford University Library, the Director of Academic Information Resources Publisher of HighWire Press, and the Publisher of the Stanford University Press. And I think you need a library just to hold all your titles and business cards, Michael.

Michael Keller:

It's a problem - I need about four or five business cards, that's true.

Hal:

Well, welcome to the Innovating at Sun show. So, why don't you start by telling us a little bit about what you do and all those different jobs.

Michael:

Well, of course, I'm gonna start off by congratulating you at Sun for your 25th anniversary, and we're proud to be one of the places that, because of our own innovative spirit, generated the possibility that Sun would become the famous and successful company that it is. With regard to the work that I'm doing, or actually, the work that I'm leading, we have at Stanford a very large research library with about 25 different branches, about nine million volumes. We circulate about three million items a year, physically, but of course, we've become a great, big digital library, as well - more about that later. Academic Information Resources refers to the part of my organization that is, essentially, the biggest but not the only academic computing organization at Stanford. We serve four of the major schools and all of the undergraduates and many of the undergraduates with various kinds of services. HighWire Press is an Internet publishing services for academic journal publishers mainly; we have 150 publishers who buy services from us to turn their physical journals into very high-featured, very high-functioning and very busy e-journal sites on the web. We physically [?] have 1,011 e-journals. You probably have heard of some of them: the Journal of the American Medical Association, the New England Journal of Medicine, the Proceedings of the National Academy of Sciences, the Journal of Biological Chemistry, British Medical Journal, Blood, Chest, Heart - I love those, Blood, Chest and Heart, the ones I try not to show up in, I'm usually the example of what not to look like as an adult. [laughter] We're also proud to be supporting virtually all of the journals of the Oxford University Press and its publications; as well, from Oxford, we're supporting their online edition, the most up-to-date edition, of the Oxford English Dictionary - we're very proud of that. Also, the final one, of course, is the Stanford University Press, which publishes roughly 130 to 150 monographs on various academic subjects each year. Those monographs, after they've been out on the marketplace in physical form for a few months, are then distributed in e-book form through a number of different channels.

Hal:

Great. So clearly, information technology plays a big role here, whether it's on the storage side or the cataloging side or the content collection and distribution side. Why don't you tell us a little bit about where you are and how you're using some of that environment today.

Michael:

Well, actually, Sun has been a perfect source of equipment for us and operating systems since the inception, really, of HighWire Press in '95. Let me spend a minute on that environment. We have about a hundred Sun servers in use there, and as Sun releases new servers with new capacities, we tend to buy them and try them. Sometimes, we try them and then buy them. [laughter] Most recently, we've taken delivery of eight x64 servers, the x2100s, which our engineers are quite excited about. They're 1U servers; they seem to provide us with a lot of power, a lot of capacity in a very small space, and we're also quite sure that they're gonna be, as most of the Sun products are, very reliable ones. So, we're happy to engage them. We're also happy that these servers are ones that allow us to use multiple different OSs, so we'll be having - some of these servers are Linux - Red Hat Linux servers. Most of them, however are Solaris servers. The big array of servers and memory devices goes to the very large size of the database that we're serving up, one that's enriched every week by thousands and thousands of e-articles. And because these are very high-impact, highly-cited journals used by lots and lots of professionals in the life sciences, the medical sciences, social sciences, we get a lot of traffic. So, a lot of your servers are being used to manage the IO to the rest of the servers in the storage array.

Hal:

High throughput, so that's good. I mean, the whole point of writing stuff is that it's not a write-only memory, you also want people to read it back at some point.

Michael:

Exactly.

Hal:

Bandwidth is good here.

Michael:

And they not only read it, but they also save it and we get reuse of these articles and essentially, remix of the ideas on a very - how to say it - it's a high-frequency basis. There's a lot going on in this space.

Hal:

I'd say probably the more cross-referencing there is and the more consumption of content there is, the deeper the links are, and of course, the more valuable the library becomes because you begin to detect what the common areas of content and the common areas of interest are.

Michael:

Precisely. And we provide a number of very high end indexing and searching services to the readers. We also have been supporting the publishers in the conversion of their [unintelligible] to digital form, so we now have most of the journals going back to volume 1 number 1 with hyperlinking of the footnotes, forward and back, to the cited references. Very high end stuff; we basically like Sun, the most innovative service in the space.

Hal:

So, you mentioned that sometimes you try things, then buy things; I understand that mostly you've been involved

in the try-and-buy and working with some of the new technology we've nicknamed Honeycomb.

Michael:

Correct. It's a - I have to qualify that a little bit, though, because it was not only a try-and-buy, it was also a head-to-head with another supplier. And we had been interacting with this other supplier for a fair period of time, but it was a black box organization, where we couldn't use APIs to get in and see how the thing was running, we couldn't affect how it was running. The try-and-buy with Honeycombs proved to us that that was a much better technology for us. We liked the fact it was fault-tolerant. The Honeycomb, in tests, and now going into operation - it's being used in our big digital repository where, really, terabytes - tens of terabytes are gonna go in there very quickly. We're very pleased with the try-and-buy opportunity, but as I said, it was a head-to-head and Sun won out in both the performance and the price categories.

Hal:

So, you talk about digital repository and, you know, we've been working with Stanford, I think, in some joint research areas for a while at how we ensure that we don't lose any of our digital content. I know we had engaged working with you on a project called LOCKSS, which, if I remember, is Lots of Copies Keeps Stuff Safe.

Michael:

Precisely.

Hal:

Our very best third-grade spelling there. [laughter]

Michael:

LOCKSS has been taken up by about 250 libraries around the world and what - the basic functionality is we have caches of content in each of these institutions where the institutions have authorization to look at the material. So, it's not a - it's a peer-to-peer situation, but it's a peer-to-peer situation in which the publishers know that the stuff is going out to the institutions. If the mother file goes down, if the network connection to the cache is bad, or whatever, the caches self-heal by checking with one another and drawing in the content that they don't have. That's basic LOCKSS. We have a new one now that we're using in cooperation with about a dozen very large for-profit publishers of academic journals, among them Elsevier, Wiley, Blackwells, Springer, in which we have a limited number of caches, in this case, 12 caches at six institutions, spread out over a lot of tectonic plate. [laughter] And the idea there is for them to deposit not just the content to which the institutions may have subscribed, but also the unsubscribed content. The idea is to provide libraries everywhere with a service at these six institutions - and soon, maybe 12 institutions - to provide a backup archive for these publishers, who are, themselves, perhaps not going to do that.

Hal:

So, you have the data out there, it's archived, it's in multiple locations. If there is accidental, or in some cases, malicious damage to it, you can go and rebuild it from the multiple copies. It's kind of like RADE on a very, very large geographic scale.

Michael:

It's a little bit different than RADE, because these instantiations [?] actually take whatever the publishers are pushing out over HTTP, which means that when the mother file goes away, what comes back is exactly what the publisher sent in, with their branding, with their functions and their features. It's not like some other archives that are, basically, intended to be [?] third party-entrusted archives, where they take in the content, then

normalize it, and the branding and some of the features just aren't available.

Hal:

So, you get out what you put into it and it looks and feels the same, which is, I guess, you know, we've historically thought of a library as a place you go - I certainly remember in my college days, the library was the place - I didn't do a whole lot of work there, it was more of a social thing, but the library was a physical place. And as you talk about digital archives and replication, library, as a notion now, becomes much more of a distributed and - I hesitate to say virtual sense - but it's much more of a place on the network, as opposed to a place on the campus.

Michael:

Well, it's actually both, and I think it's too bad that Princeton guys only go to a library for socialization. [laughter] I'll be sure to tell my friend, the librarian at Princeton, to do something about that. The truth, though, is... [overlap]

Hal:

There's probably still a black mark next to my name there in the registration book. I wouldn't be surprised.

Michael:

You can make this good by sending them a nice, big check. [laughter] The truth, though, is we have to think about libraries in one large sense and realize it has a couple of different manifestations. The one large sense is that these are places where people choose information to be managed, they choose information to be brought into a community to be used by members of that community. Whether the information artifact is a physical one or a virtual one, now is the game. And how we deliver services, based on that [unintelligible] new paradigm is a complicated and interesting matter.

Hal:

That point about content coming into a community, physical or virtual community, to be used, if you want to think of that as a very Web 2.0 view of the library, that there is publication of content and then there's adoption of it or consumption of it, you know, by people, you know, no matter where they happen to be. And the things you talked about in terms of being able to link things together, being able to get the historical volumes, really drives this notion of finding all the related content, or all the things that would be in the same taxonomy, or same conceptual index of, you know, those same ideas.

Michael:

You're exactly right. And actually, we've been building out that - what became known as the Web 2.0 library from the period '95. The third thing, though, is to realize that we are also building out the features that will allow Web 3.0 to deliver automatically to our readers what they have found interesting in terms of subject or language or author in previous uses of the information set.

Hal:

And that, I assume, will allow you to extend, not just in terms of the types of content there, but also, you start to look at now how people actually learn, whether they actually go through the content you have. And, you know, inside of Sun, this is one of our, you know, large points of discussion around training and skill sets, which is, with all the resources we have available to us electronically, how do we keep track of what actually worked? How did someone step through all the different resources and then we go and create a learning path out of, literally, a safari map through all these different - the wild zones of online content.

Michael:

Yeah, exactly right. That approach, that examining the track and the trail, is, of course, one very good approach to understand how people learn and how people might learn better. There are a bunch of other tools that go to assessment, as well; we are using them in our digital language lab. It's a field that's opening and it's a line [?] for more possibilities, more methodologies, of interest and of use across a very much wider range of cultural styles.

Hal:

So, what's next on the farm?

Michael:

Well, we're building a bookless engineering library. It's in planning right now, and it's made possible by the fact that the literature of engineering has gone almost completely digital. Eventually, it will be a totally bookless library, and opening it, we'll probably have about 20 percent of what we now have on the shelves in the engineering library. And as we think about that, we're now thinking about, really and truly, that conundrum that you presented, that new paradigm of the physical and the virtual library - the cybrary. But imagine now what it is that librarians will do when the literature that they're most interested in and the interactions with that literature their patrons will have, are working with literature that's entirely virtual. It's a big challenge, but I think it's going to prove that we can, as information professionals, serve the needs and interests of the searchers and teachers and even students better because we're better able to understand what they're doing with the stuff, what they're interested in, and how we might deliver it to them in ways that they find most useful.

Hal:

And certainly, as you do that, you'll also be able to pull ideas from disparate courses of study or disparate industries and start to glue them together and find new ways of innovating.

Michael:

You're exactly right. It does - interdisciplinarity does drive innovation, as does the speed of reporting and then remixing and reusing ideas and expressions, no question about it.

Hal:

And certainly our hope that the pace doesn't slow down and the innovation continues. As you mentioned at the outset, Sun Microsystems, as an outgrowth of work that Andy and others did at Stanford, you know, is a result of that spirit of innovation and that, you know, emphasis on research and just doing things a little bit differently. And here we are, 25 years later, you know, still talking about it.

Michael:

And still imagining better ways of doing our work. It's really a great environment.

Hal:

We certainly look forward to continuing to work with you, you know, in the immediate future, and hopefully, over the next 25 years, as well.

Michael:

Great, I'm looking forward to it, too.

Hal:

Well, thanks for being a guest on the show. Again, you have been listening to the Innovating at Sun on Podcast,

<http://blogs.sun.com/innovation>

and I'm your host, Hal Stern.

Announcer You've been listening to Innovating@Sun. Join us next time for the latest in innovation from Sun Microsystems. Only on the Sun Microsystems Podcast Network.