December 2002

ATG Interview with John Sack

Margaret Landesman
University of Utah, margaret.landesman@library.utal.edu

Follow this and additional works at: https://docs.lib.purdue.edu/atg

Part of the Library and Information Science Commons

Recommended Citation
Landesman, Margaret (2002) "ATG Interview with John Sack," Against the Grain: Vol. 14: Iss. 6, Article 15.
DOI: https://doi.org/10.7771/2380-176X.3947

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
ATG Interview with John Sack

Director of HighWire Press, Stanford University Libraries

by Margaret Landesman (Head, Collection Development, Marriott Library, 295 S. 1500 E. University of Utah, Salt Lake City, UT 84112; Phone: 801-581-7741) <margaret.landesman@library.utah.edu>

John Sack is the Director of HighWire Press, a Department of the Stanford University Libraries, which assists with the publication of 334 journals, mostly in the sciences...

Author’s note to readers who want to know the bottom line up front:

Particularly compelling among the things that I learned while talking to John are:

1) HighWire Press is the publishing partner of 43 of the titles ISI lists as most highly cited and of 77 of the top 300 titles.

2) HighWire is also responsible for putting the Oxford English Dictionary online.

3) HighWire encourages its publishers to make back issues freely available after a period of time — six months, a year, two years — they are comfortable with. About half of them have done this, creating what Stanford’s librarians believe to be the single largest free repository of scientific literature, with over 465,000 free full-text articles, adding about 10,000 more each month.

4) HighWire’s mission was originally and continues to be to ensure the strength of scientific societies as publishers. It is not about putting 300 journals online. It is to make sure the societies are successful in their electronic publishing programs.

5) HighWire isn’t a traditional aggregator, since it doesn’t own any rights to any of the content. It improves the usability of the scientific literature for scientists by allowing publishers to choose common tools and interfaces for searching and reading articles, and by providing a common repository — the HighWire Portal at http://HighWire.stanford.edu — which provides single-point searching and alerting for all of Medline plus all the full-text that HighWire hosts. But HighWire doesn’t compel publishers to adopt common license terms; it brings the societies and publishers together twice a year to see if they can find common approaches that will work for their publishing programs as well as the institutional subscribers and their memberships.

ATG: HighWire Press and the OED — it’s not a journal and it’s not science — how come it’s HighWire?

JS: When Oxford put out the invitation to tender for the OED, there was no question that HighWire was going to respond. HighWire has many staff who were undergrad or grad students in English and the Humanities, so this was a resource many of us were familiar with, and, in at least my case, in love with. If you think the OED is “fun to read,” you probably would feel right at home at HighWire.

I had two copies of the Compact Edition at home — remember when the Book of the Month Club offered it for $75? It’s a fundamental resource; and it’s easier to use online, though perhaps not as pleasant. (Though those of us who bought the Compact Edition in the 1970s probably now find we can’t read the micro-reduced type even with our glasses and the magnifying glass.) HighWire did try hard to imitate the “serendipity effect” of opening a dictionary — to make it so that when you look up a word, you notice the words that are around it. The OED staff at Oxford thought hard about this in their user interface design and we think it is a good way of browsing and “discovering” related entries in a very complex reference work.

I had worked on this sort of issue before. When Stanford built its online catalog in the mid-80s, keeping the “good part” of the online card catalog and the stack shelves — the way in which related information is found together — in moving the catalog online was a fundamental part of our goal. We wanted, to the extent that we could, to imitate the “serendipity effect” of finding a book on the shelf — or not finding it — and browsing nearby titles.

HighWire does host mostly science journals — and the fundamental technical tools to do scientific text and other sorts of very complicated text such as the OED are not different. The user interface, of course, is entirely different — a dictionary is a special type of reference work, and what HighWire had done before was neither reference nor dictionaries. And the OED is a very special type of dictionary.

ATG: How does HighWire come to have such a high proportion of highly cited titles?

JS: I was impressed by the numbers of highly cited titles noted on HighWire’s Web page and asked about that. There are all sorts of different ways of using numbers, so one has to treat all such statements with some care, but it is true that about half of the most frequently cited journals are HighWire-hosted journals. No other online service can say that, because most societies, of course, even largish societies, don’t have that many journals, and big commercial publishers have many more journals, but none of them have half of the most cited journals. So it is unlikely that any other service can put together such an impressive collection of full-text “under one roof.”

What does this show? The best journals are society journals and journals from non-profit publishers (such as PNAS, Cold Spring Harbor Labs, Rockefeller University Press, etc.). Or an awful lot of them are.

HighWire early on began its work with the most frequently cited titles in science: the Journal of Biological Chemistry, Science, and PNAS. As it turned out, the editors of these journals were on the publications committees and editorial boards of other highly cited journals, and they would talk to each other about the work they were doing with HighWire. That “word of mouth” meant that HighWire didn’t have to toot its own horn, but could grow through doing good work.

ATG: What is your academic background and how did you come to be involved in founding HighWire?

JS: My background is in English Literature. I was a graduate student at Stanford in the 70s and became fascinated with trying to understand how a writer could have written something centuries away from the time in which the reader is reading it — and still be understood. My interest was in modern literature, but I had to do coursework in Old English and Middle English. And what I found odd about Beowulf and Chaucer was that after all these centuries since they were written, I can still understand them.

I wanted a metaphor I could use to study this phenomenon and came up with the idea of looking at how librarians have stored (cataloged) information so that people in other times and other places can retrieve it — how they store information in library catalogs and in storage and retrieval systems.

This took me into SPIRES, Stanford’s online catalog. I was hired on at Stanford as a documentation writer (there being no jobs for English professors then!) and ended up managing the SPIRES Development Group. Through SPIRES, I became interested in large bibliographic databases and went on to work with RLIN and RLG as a director working for Ed Shaw in the 1980s. Stanford in the mid-90s combined its libraries and its technology resources into a unitated “LAIR,” Stanford University’s “Library & Academic Information Resources.” The libraries and the Stanford Data Center, academic computing and networking all were a part of this group. I was Director of the Stanford Data Center, which provided large-scale database support (such as a library online catalog would require) and mainframe/server support.

When the Stanford University Libraries hired Mike Keller as library director, the directors of LAIR were involved in the search. The directors thought a lot about how to identify a library director who would turn out to be good at working with the rest of us technologists in this unusual libraries-and-technology organization. We decided to construct an exercise — a joint problem-solving session — in which the candidates and the directors could see how they worked together. We chose the serials crisis as the topic. It was in the session with Mike Keller that the idea of HighWire was really invented — that a university should work with many high-
impact science publications to develop technologies and tools to distribute content using the Web — though it wasn’t until later that we would know this.

About a year later, after Mike made a presentation on the serials problem to the Stanford University Senate, Bob Simoni, a Stanford faculty member and the deputy editor of JBC approached him. This was at a time when the JBC was “only” about 1000 pages/week, and the advanced technology was a CD-ROM; a new CD-ROM was filled about every 6-8 issues. Maybe — we thought — maybe we should actually do it online on the Web. In the general university research environment, the Web was a new thing, and “gopher” servers were still prevalent.

**ATG:** How did you go about creating HighWire?

**JS:** We didn’t look at other implementations very much. There were only a few in early 1995 — and nothing that we believed really “sang” to scientists — i.e., that they recognized as just what was needed to find and read the literature. Moside had just been born, and Adobe’s Acrobat PDF format was still very new.

There were people putting print journals online. With their tools, you could look at images of printed pages and browse print journals. And there were some clever interfaces which featured nice page turning and so on.

But our insight was fundamentally different from putting print online. We saw a science journal as a database. What readers seemed to want was not a journal per se, it was the information locked in the collection of articles. We saw ourselves as giving people ways to get the information out of the database and out of articles.

We were not thinking so much of the people who wanted to read the current issue of a journal. We were thinking of those looking for information, who needed a database with a search engine. And we weren’t interested in the print-format pages (at the time) — we were interested in the searching and scanning and reading.

At the time, we were onto a couple of things, maybe three, that others had perhaps not yet fully realized.

Our first insight was that the literature could be adequately represented from the scientists’ point of view using a Web browser. We knew scientists didn’t want to deal with special software applications to read a journal. We knew they didn’t care how pretty the pages looked — they just wanted the information to be accurate. It took a lot of work — we had to build little GIFs for all of the Greek letters and other symbols — but we were able to utilize the Web browser’s limited capacities creatively to do everything that was needed to represent scientific information.

Second, because we were working with scientific societies, our focus was more on how you distribute information than on how you collect money for it. Any publication system has to do both eventually, but we were less interested in where you put the tollbooth.

Third, we were very aware that we were building a big database, with an overwhelming amount of content. The Journal of Biological Chemistry was our first journal and JBC was the second biggest journal of them all. It was really weird: and once we knew some body printed out the new articles and mailed them to people. Working with JBC helped us to generate all the basic concepts for HighWire.

Other early efforts from other pioneer groups in the online journal space started with small journals, which had to deal with relatively trivial amounts of content. They focused on handcrafting things. We, on the other hand, knew we would get 120 articles a week — a typical JBC issue — and everything had to run through programs. We couldn’t have people typing things in.

A downside to our startup approach — working with scientific societies who typically publish one very high impact journal — is that you see things one journal at a time. Big commercial publishers understood they would have hundreds of journals and they planned for it. We didn’t. We looked at how each journal was supposed to work for its own community. For example, as we worked on Pediatrics, we thought about how the needs of pediatricians might differ from those of molecular biologists.

**ATG:** What else does HighWire do for publishers and libraries?

**JS:** As a part of HighWire’s mission, we try to help them to share information with each other. Twice a year we bring them all together. We hope this speeds the evolutionary process so that bad ideas eventually get weeded or described by one publisher without other publishers having to “repeat the experiment.” And, of course, that good ideas get propagated faster.

We have also been working on our new HighWire portal to pull together for libraries much more of the subscription and business information in one place. We have tried to use it both ways — so that our publishers retain control of their policies and business models while libraries, though they still have to deal with all these different publishers for subscriptions, at least don’t have to go to 150 different places to change an IP address.

But in the end, it is the publisher who pays our bills. We see our role as providing software layers that bring information together, but we want to make sure we don’t get in the way of the relationship between the library and the publisher.

**ATG:** I was surprised to see such emphasis on “pay per view” on HighWire’s Web page.

**JS:** A year and a half or two years ago, there was a movement with the Public Library of Science to make all literature free short after its publication. Cold Spring Harbor Labs pulled together a group of scientists and publishers to talk about this, and HighWire was a part of that discussion. One of the things I learned at that meeting was that “free” wasn’t the requirement, “accessible” — as in “I can get it on my workstation” — was the requirement. It would be all right, some scientists said, if there were an inexpensive “pay per view” option available for information that wasn’t free. Not $30. Maybe $5 — $10. That would be good enough. Not quite as good as free, perhaps, but good enough. As long as the information is accessible, that is what is important.

For generations, everything has been based on subscriptions. To only pay for information when you need it is a totally different model for librarians and for publishers. About two years of HighWire’s journals now offer “pay per view” to make content more accessible.

**ATG:** Isn’t HighWire also known for the free back content it offers?

**JS:** Many HighWire publishers do make back content free, most of these after a few months. A huge amount of the information in HighWire is free, about 60% of the total. Not quite half of the journals have free back issues, but as the biggest journals (e.g., JBC and PNAS) do, it’s more than half of the total content is free. Many also offer “toll free linking,” so that if a researcher is working in a journal to which the institution subscribes and follows a citation reference to an article in a journal to which the institution has no subscription, the individual can nonetheless use the cited articles in the nonsubscribed journals for free.

Many of our publishers are very active just now with back conversions — it’s one of the big trends we are seeing in the new century. Scientific societies have started to realize that as scientists become more and more used to using stuff online, many are reluctant to go back into the paper. They just won’t read what they can’t get online. Eleven journals that HighWire hosts have now completed significant projects.

Forty-two journals have projects in progress. About another dozen have projects moving along in discussion phases. These projects are mostly putting back content only back to 1990 or 1980, but several societies are going back farther and a few are planning over the long run to have their entire content for a journal from volume 1 online (Science and PNAS already have this scope of content online, through JSTOR).

There is interesting data that shows that older stuff does get used, but the decay curve is stunningly rapid, within six months the use of an article drops by an order of magnitude. Any author would feel that they are a pop songwriter or something like that. But as far as we’ve looked, over years of data, the use doesn’t go to zero. It continues at a low level. This shows the power of a search engine — if you have articles online, they will get used.

**ATG:** How do you feel about the LOCKSS (Lots of Copies Keeps Stuff Safe) local archival solution being developed at Stanford?

**JS:** I have been fascinated by how unusual LOCKSS is as an approach. It makes sense, and it is very different from other approaches. I’ve concluded that it is one of the right approaches. We need multiple approaches — different protocols to be used by different groups — so that if it turns out there is a fundamental flaw in one approach, it won’t wipe out our access to the information.

There’s the question of who takes responsibility for what. If libraries build redundant archives and then publishers put archives in other places, that combination is a good one. No one continued on page 64
John Sack Interview
from page 62

Group on its own is enough. We need a sort of
"biodiversity" model, with several groups tak-
ing different approaches.

ATG: On a lighter note, I wondered where
HighWire Press is located?

JS: HighWire is in the Stanford Industrial
Park, about half a mile from the Stanford
Library and the center of campus. It used to
be that my office was in the library and the rest
of the staff was in trailers right next to it, but the
university didn't like having to continually add
trailers as we grew. The University had a build-
ing available in the area right off campus (across
the street from Hewlett-Packard's main building),
and HighWire leased it. It's a good space
and centrally located, with easy access to the
campus. But the best thing is that we got the
University to agree that we wouldn't have to
have parking permits. As permits cost $200-
400 dollars a year, this was a very popular deci-
sion among HighWire staff.

ATG: The photos on HighWire's Staff Di-
rectory seem to show lots of dogs and babies
on the payroll.

JS: Well, we have about a dozen dogs in
during the day. It's definitely a dog friendly
and child friendly place. It's good when kids are there
— there's usually only one at a time. Many of
us don't have kids of our own, so having kids
around is an invigoration thing. It reminds you
of the rest of the world. And dogs are fascinat-
ing because they are dogs — running around
doing doggy sorts of things. Kids and dogs de-
stress you pretty rapidly.

ATG: Perhaps we could end with some
ideas about how you see the future of HighWire
— what do you see happening next?

JS: What we see happening is that in a few
years, people will be doing most of their litera-
ture discovery work in just a few places. There
has been quite a consolidation of general-pur-
pose Internet-search sites and we see this trend
continuing on into the science-search domain.
People form habits as to where they work. If
HighWire isn't one of those places, our pub-
lishers' content could be ignored when research-
ers are doing "discover" (vs. "known item")
types of searches.

Researchers will use multiple approaches to
literature discovery — and that is a good thing.
But they will use a small number of different
approaches — choosing just a few from among
the most powerful and useful tools and content.
Our goal is to make HighWire's portal at
http://highwire.stanford.edu one of the choices
on their short list, to be sure our publishers' con-
ten is highly visible.

ATG Interview with Victoria Reich

Director (and founder) of the LOCKSS program <http://lockss.stanford.edu>

Addressing the Challenges of e-Journals, February 28, 2002

by Margaret Landesman (Head, Collection Development, Marriott Library, 295 S. 1500 E. University of Utah,
Salt Lake City, UT 84112; Phone: 801-581-7741) <mlandesm@library.utah.edu>

As Vicky and I talked, we discussed her
various activities in more or less chronologi-
al order. Later on, Vicky said of the work
we'd been discussing, "but that's not my true
love of the moment."

There's not much to say to that, except to ask:

ATG: And what is your true love of the
moment?

VR: That would be LOCKSS, Lots of
Copies Keep Stuff Safe.

ATG: Why this fondness for LOCKSS?

VR: As with HighWire back in 1995, we
have a new idea about how to solve a difficult
problem, this time e-journal archiving and
preservation. This is a critical issue for both schol-
arly publishers and libraries. So far, the
LOCKSS technology is promising. We're in
the early days, developing the idea and helping it
come to life. I've been given an opportunity to
work with many international librarians and
publishers. Meeting new colleagues and helping
to solve an important problem is exciting and fun.

ATG: How did the idea of LOCKSS come
about?

VR: About three years ago, while hiking
with a friend, I was voicing frustration with the
lack of progress around digital archiving and
preservation. HighWire had built an excellent
electronic journal online environment. We had
made it easier for readers to use and find the
literature. However, even three years ago, the
authoritative version of most titles was not the
paper version, it was the electronic version. The
paper edition of many titles provides a subset of
the whole journal, which is digital.

This is causing a serious problem. Librar-
ian's aren't giving up paper copies because they
worry about archiving — as they should. But
keeping the paper is no longer archiving, be-
cause paper isn't the definitive edition. So, li-
brarians are spending money on the print, which
is no longer the authoritative version. Even
when they purchase the electronic, in addition
to or instead of the print, they can't take cus-
tody of the content. In the current electronic
journal world, librarians, through little fault of
their own, have abdicated building local collec-
tions. They are unable to take custody of mate-
rrial, to guarantee access to important material
for both their current and future communities.

And there is important literature at risk here.
For example, take Pediatrics, the journal of the
American Association of Pediatrics. Pediat-
rics has unique content in its paper edition and
almost since day one, substantial unique con-
tent in the electronic. But libraries aren't cap-
turing the electronic part.

ATG: Ok, my friend responded, tell me
about how libraries work. How do they do it
with paper copies?

VR: With paper, there are often many copies
of any particular journal scattered around the
world. Libraries know where some of the cop-
ies are, though it's difficult to find all of them.
We talked about people who go to libraries
with razors and blades, and the ways in which
some paper copies might get destroyed. But,
because the system is "redundant," though we can't
find all of the copies, we can easily find some of
them and deliver them to authorized readers.
Could there be something like that for the
electronic world; would it be possible to build a
system that allows libraries to take ownership of
the journals and to have a similar level of
security in archiving?

ATG: So how did you go about building
LOCKSS?

VR: With the support of [Stanford Uni-
versity Librarian] Michael Keller, we received
funding from Michael Leck, then of the Na-
tional Science Foundation, to build a first ver-
sion of the system. AAS ScienceOnline gave
us permission to use their content for testing. We
distributed ScienceOnline content to six part-
er libraries (Berkeley, Columbia, Harvard,
LANL, Stanford and Tennessee). Each library
brought up a persistent LOCKSS cache and cached
the ScienceOnline test content. We deliberately
damaged the content files. Some-
times content files just became damaged due to
disks failures or major fires in New Mexico, or
human error. As planned, the machines talked
to each other and repaired the damage. The ma-
chine collected new issues as they were published.
Over the course of that year, the scheme appeared
viable and the specific tests worked.

We then requested funds to continue testing
from the Andrew W. Mellon Foundation. Fortu-
nately Mellon staff agreed that LOCKSS
seemed like a promising idea and with their
support we've been running a beta test. The
beta test has over 50 participating libraries, about
half of which are in the U.S. We're using PNAS
Online, ScienceOnline, bjm.com, and the JBC
Online as test content. Many publishers have been
terribly generous and supportive; we
continued on page 65

64 Against the Grain / December 2002 - January 2003 <http://www.against-the-grain.com>