

COMPLETE COVERAGE OF LIBRARY INFORMATION TECHNOLOGY

# COMPUTERS IN LIBRARIES

THE  
LATEST ON  
STANDARDS

Volume 27, Number 4  
ISSN 1042-7915

APRIL 2007

## How SUSHI<sup>6</sup> Harvests Statistics

Learn About  
a Dozen  
Standards<sup>11</sup>

Format  
Wars<sup>24</sup>

COMPUTERS IN LIBRARIES 2007  
CONFERENCE EXHIBIT DIRECTORY  
PAGES 31-33

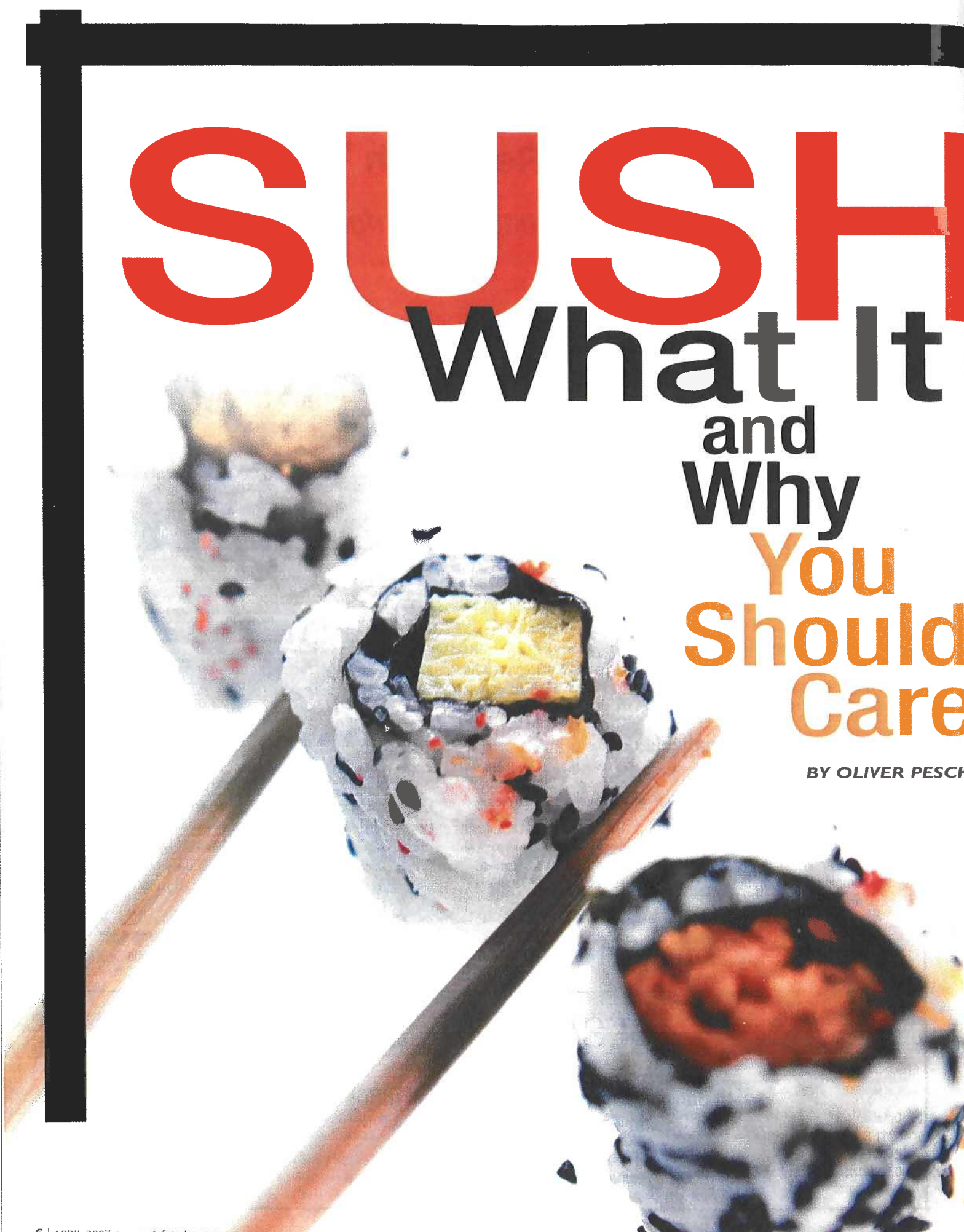
# T SUSHI

What It

and  
Why

You  
Should  
Care

BY OLIVER PESCH



In summer 2005, a small group<sup>1</sup> consisting of two librarians (Tim Jewell and Adam Chandler), an ILS vendor (Ted Fons), and a representative from NISO, the National Information Standards Organization (yours truly) had an informal meeting in the lobby of the Hyatt McCormick hotel in Chicago. The occasion was ALA's annual conference, and the purpose of the get-together was to continue a conversation that Adam and Ted had started the previous fall about seeking a better way to collect usage statistics. At the time, none of us would have predicted that out of this casual meeting would come not only a new NISO standard (Z39.93), completed in record time, but also a new term for our industry.

Two years ago, if you were to ask librarians about SUSHI, they'd think you were inviting them to lunch at a Japanese restaurant. Ask those same librarians today and there's a good chance they'd assume you're talking about usage statistics. So how did we go from a meeting over *caffe lattes* to one of the most recognizable standards in NISO's long history? What problem were we solving? What is SUSHI? How does it work?

### Harvesting and Normalizing a Variety of Raw Statistics

First, let's start with what SUSHI stands for—standardized usage statistics harvesting initiative. From this, you can see that we're talking about usage statistics and harvesting. Why are usage statistics important?

With the success of the online format for journals (and now books), librarians are far more interested in measuring the use of these resources. In the days when a library was mainly about print collections, staffers used circulation and shelving statistics as part of the collection management and decision-making processes. Librarians then (as they are now) were responsible for managing their budgets effectively. One measure of success was to ensure that they were spending money on needed resources, i.e., those that were being used. This is just as true with online collections as it is with print.

A metric that's now becoming an important input in the online-collection decision-making process is "cost per use." This is simply the result of dividing the annual cost of the journal by the number of times the full text was used. (For some who may be concerned that cost per use is the sole element that goes into decision making, librarians realize that collection management involves many other factors, such as qualitative measures like the stature of the journal and its impact on institutional research.) So far, so good. But where does the harvesting problem come in?

When the world was print-based, virtually all activity centered on the library as a physical place. Therefore, since staffers were actually handling the materials, they were able to control the monitoring and collection of statistics as part of day-to-day routines. In the online world, the library has become distributed across the Web, with end users accessing content on dozens or even hundreds of publisher and database vendor sites. This means that instead of the librarians, it's the publishers and database vendors who are keeping track of what's being used. As a result, usage data is scattered across the Web, hidden behind dozens of administrative modules. Keeping track of where to go to collect the usage numbers is a significant problem that's growing worse as more and more content becomes available online.

But it gets more difficult. Finding the usage data was only the tip of the iceberg. Once it's discovered, you have to navigate a variety of user interfaces and pull the reports, only to find the formats don't really match. So if it's comparable usage data you want, you have more work to do to normalize it. But aren't there standards out there already that address this?



"The promise of SUSHI is that, with wide adoption, it will drastically reduce the time spent collecting usage data and will greatly increase that data's quality and consistency."



