

INFORMATION STANDARDS QUARTERLY

FALL 2011 | VOL 23 | ISSUE 4 | ISSN 1041-0031

TOPIC: STANDARDS IMPLEMENTATION BENEFITS

NISO CIRCULATION
INTERCHANGE PROTOCOL (NCIP)
IMPLEMENTATION COST SAVINGS

SCHOLARLY iQ AND SUSHI
IMPLEMENTATION

BENEFITS FROM USING SERU:
A SHARED ELECTRONIC
RESOURCE UNDERSTANDING

INNOVATION AND
STANDARDIZATION

NEW NISO INITIATIVES:
DIGITAL ANNOTATION SHARING;
OPEN DISCOVERY

Gary Van
OverborgJohn
MilliganMichael
Lee

Scholarly iQ and SUSHI: A Case Study

GARY VAN OVERBORG, JOHN MILLIGAN, AND MICHAEL LEE

Scholarly iQ serves over 30 publishers to provide timely and accurate COUNTER-compliant reports to their thousands of subscribing institutions. We have been a leader in COUNTER compliance reporting since 2002. Our publishers enjoy the benefits of an actionable set of key performance indicators plus the power and flexibility of a web analytics and optimization engine which fully integrates with our COUNTER reporting engine and their offline data.

Service-oriented architectures are surging in popularity. Web services, for both publishing and consumption, are working their way across virtually all business interactions. The ability to make data available on demand through standards-based interfaces has transformed the way that organizations interact. *The Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol* standard (ANSI/NISO Z39.93) has brought this transformational technology to the COUNTER reporting space, and throughout its development and adoption, Scholarly iQ has been on the forefront of the effort to take SUSHI mainstream.

Being an early adopter is not without pain, however, and it is the responsibility of early trailblazers to lay guideposts along the way. This case study is a collection of these guideposts. Our objective is to identify some of the struggles, successes, and observations that we have seen along the path and to share those with the community. As with any standard, SUSHI will continue to increase in relevance and utility as its level of adoption expands. It is our hope that our experiences will provide additional direction and motivation to those that are contemplating the pursuit of this innovative protocol.

In this case study, we will first examine some of the challenges that we experienced in our implementation of the SUSHI specification. Then we will discuss some of our observations regarding the current trends in SUSHI adoption and usage. Finally, we look forward to see what SUSHI has on the horizon, and how Scholarly iQ will play a role in that future. We are pleased that you will be making this journey with us.

Being an early adopter is not without pain, however, and it is the responsibility of early trailblazers to *lay guideposts* along the way. This case study is a collection of these guideposts. Our objective is to identify some of the struggles, successes, and observations that we have seen along the path and to share those with the community.



■ CHALLENGES

Scholarly iQ faced a number of challenges as we developed our SUSHI implementation to enable our customers to harvest their usage data electronically in an XML format using an automated retrieval process. In addition to development challenges, we also faced issues when testing the service externally with clients and vendors. These tests exposed other difficulties including compatibility and data related issues.

Development Challenges and Resolutions

One of the main challenges that we faced was figuring out where to begin. Since at the time, SUSHI was a relatively new protocol with little or no information available on any past development, it was difficult to get an understanding of what the first steps should be. We were provided with a SUSHI protocol standards document that included useful information on the schema (XSD) and report objects, however we had no true direction on how to begin. We overcame this challenge by diving in and getting involved with the NISO organization, analyzing how our data translated into the schema provided, and how these pieces fit together to meet the objective of exchanging usage data in a more efficient manner.

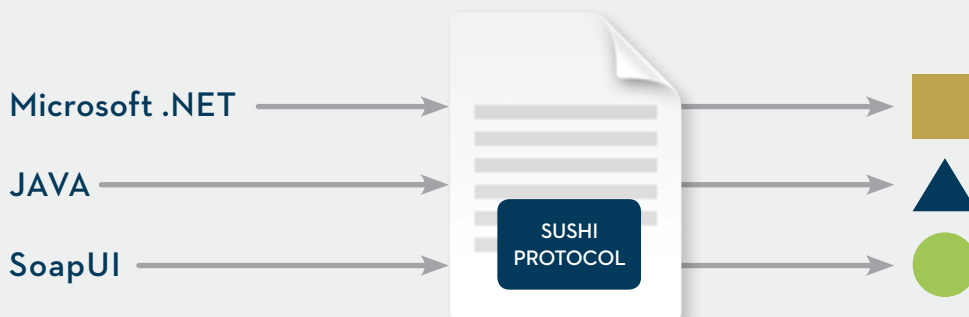
All of the elements needed to create a SUSHI Web Service were reviewed including:

- » Core SUSHI XML schema (XSD)
- » Web Service Description Language Document (WSDL)
- » Report Request Diagram
- » Report Response Diagram

In addition to these items, we also consulted the COUNTER Release 3 schema, WSDL, and report diagrams. We were then able to create a flow diagram illustrating where each of these elements fit into our overall development process. This allowed us to design the business and data layers that are critical for a low maintenance, yet highly scalable architecture. We used the Microsoft .NET technology stack for this implementation and within a few months, we had a fully operational SUSHI web server and service that could support multiple vendors.

Subsequent to starting the development, more information became available from developers and vendors that contributed information to the SUSHI community but most of this information was related to the input. Finding samples of the output (i.e. the Report Responses) was still difficult to come by during the early stages of SUSHI implementation. To assist future development efforts, Scholarly iQ made contributions to the community by providing various sample SUSHI reports that allowed other developers to see how to structure their Report Response.

CONTINUED »



CHALLENGE: The SUSHI protocol document may be subject to interpretation. In this case, there were critical differences in the way each tool processed requests.

SOLUTION: Working closely with vendors that provide client services was essential to ensure that we were able to handle a variety of SUSHI clients appropriately.

Finding Client Applications and Vendors to Test the Service

Now that we had what we felt was a fully operational and internally tested SUSHI web service, we needed to test the waters externally. One of the only tools that was available at the time to harvest usage data from a SUSHI web service was from the Euclid Project. This CGI-based tool helped us to harvest usage with a raw XML report response. As we began to look for vendors that were interested in testing their SUSHI client, we were contacted by Innovative Interfaces, Inc. They had just completed their SUSHI-based client and needed a service to test it against. This was definitely to our mutual advantage and though we had high hopes that the initial testing would be successful, it wasn't. We were plagued by permissions problems and other issues pertaining to calling conventions such as WSDL vs. ASMX. Once we resolved these issues, Innovative Interfaces was able to successfully harvest usage data from our SUSHI web service.

Compatibility Testing with Various Clients

After successful testing with Innovative Interfaces, we had the opportunity to test a new open source SUSHI client with Serials Solutions. It was then that we found out that the SUSHI protocol document may be subject to interpretation. In this case, there were critical differences in the way each tool processed requests. Working closely with vendors that provide client services was essential to ensure that we were able to handle a variety of SUSHI clients appropriately. This testing gave us the opportunity to enhance our services to meet various harvesting requests, while at the same time maintaining conformance to the standard SUSHI and COUNTER 3 protocols.

XML Restricted Characters in the Data

The Scholarly iQ SUSHI web service was put into operation in June of 2009. The challenge now was to harden the service in a production environment. While the service itself seemed stable, data-related issues started to appear that we did not consider during initial testing. One of our customers who had signed up with our service and started harvesting usage data reported that they were receiving an error in their output relating to an XML parsing issue at a particular line. This in turn, caused the Report Response object to fail; however it did not contain a standard SUSHI exception code, but rather a database exception that was not clear to the problem at hand. After thorough troubleshooting, we identified the issue to be related to response data that contained special or “restricted characters,” which violated the structure of a well-formed XML document as defined by the parser.

For example, a journal title may contain an ampersand (“&”) instead of the word “and”. This may be fine when reading data to be displayed in a report or web form, but it will cause XML parsing issues which results in an “illegal characters in path” exception.

Once we identified these cases, the solution was simple. We created a method within the web service using regular expressions that would check for these “special characters” and have them replaced with their “reference” equivalents on the data side prior to parsing. Therefore, scrubbing the data prior to sending a response resolved the issue.

Performance

As with any data service, performance and throughput are always of primary importance. When we initially designed our SUSHI web service, we were concerned about how the volume of data returned to the user during web service consumption would impact performance. Over the years, we have structured our data by horizontally partitioning the data across vendors and reporting years. This served us well in the long run as we were able to build the SUSHI web service to retrieve data from specific data sources instead of needlessly sifting through multitudes of records contained in one data repository for all years and vendors.

Changing of Historical Data

Inevitably, usage data will need to be reprocessed for various reasons. Sometimes multiple months of historical data will need to be re-processed. When this happens, institutions and consortia must be notified of the impacted date range so that they can retrieve the updated statistics via SUSHI. To assist with this effort, we developed thorough logging mechanisms that track who is using our SUSHI web service and how the service is being used (i.e., which reports are being downloaded, what date ranges are being used, and what is the frequency of these downloads). We are then able to verify that all impacted accounts are able to re-pull the corrected usage data and, if needed, “push” the updated statistics.

CONTINUED »

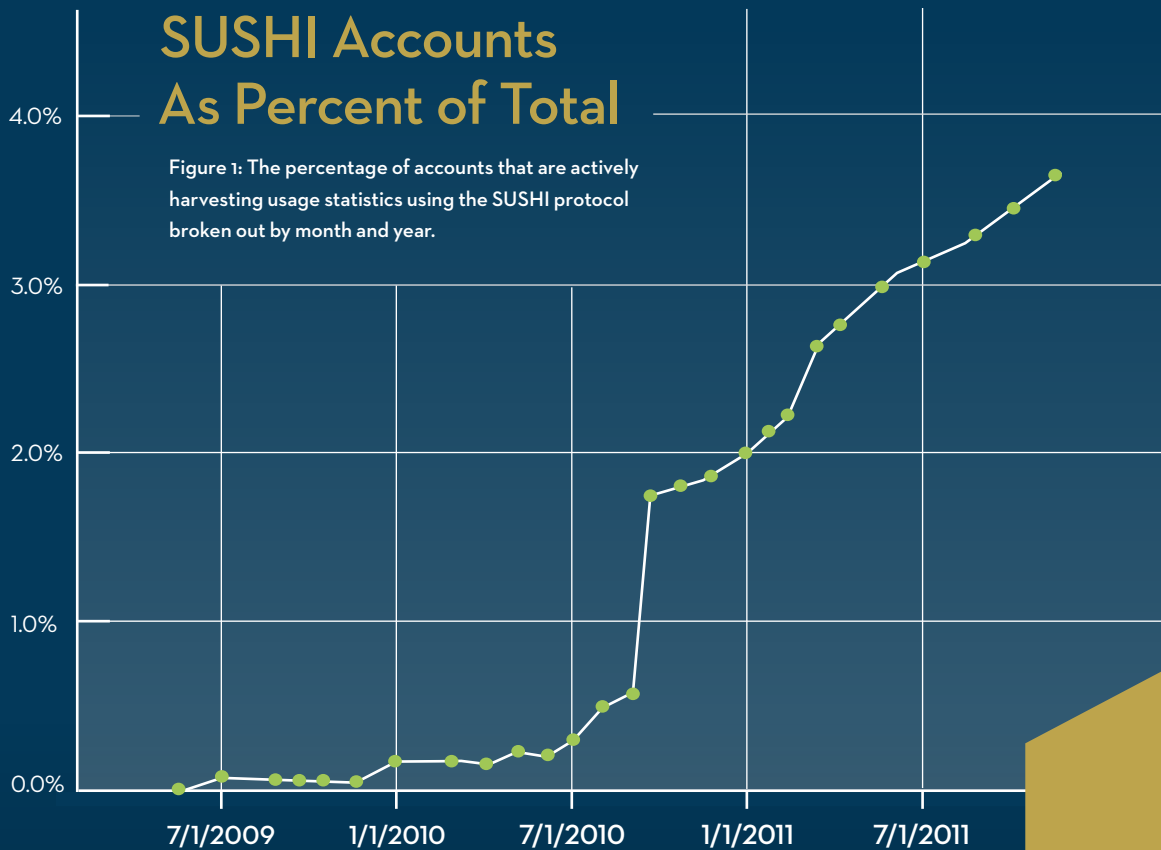


CHALLENGE: A journal title may contain an ampersand (“&”) instead of the word “and”. This may be fine when reading data to be displayed in a report or web form, but it will cause XML parsing issues which results in an “illegal characters in path” exception.

SOLUTION: Create a method within the web service using regular expressions that would check for these “special characters” and have them replaced with their “reference” equivalents on the data side prior to parsing.

SUSHI Accounts As Percent of Total

Figure 1: The percentage of accounts that are actively harvesting usage statistics using the SUSHI protocol broken out by month and year.



TRENDS

Since the SUSHI implementation is still somewhat new, it is difficult to make general statements about specific trends because there is not enough data available. While our initial analysis indicated that we were trending toward an increase in SUSHI harvesting, with a corresponding decrease in traditional report harvesting, there is simply not enough data yet to be able to say definitively that there is a statistically significant relationship in the time series analysis. However we will continue to monitor these trends and report our findings via our website and twitter.

What we can say definitively at this point in time is that while the current percentage of SUSHI harvesting accounts versus traditional harvesting accounts is small at 3.5%, it is increasing rapidly. In fact, it is increasing so rapidly that Scholarly iQ has invested in augmenting our application infrastructure with new tools that our clients can use to facilitate the sign-up process of creating new harvesting accounts that allow access to our service. Figure 1 represents the percentage of accounts that are actively harvesting usage statistics using the SUSHI protocol broken out by month and year.

While the current percentage of SUSHI harvesting accounts versus traditional harvesting accounts is small at 3.5%, it is increasing rapidly.



LOOKING TO THE FUTURE

As the COUNTER standards continue to evolve with the upcoming Release 4, so too will the SUSHI standard in order to adapt to these latest initiatives. Scholarly iQ is already in the planning phases to support the new versions and we strongly recommend vendors providing SUSHI service do the same as early as possible.

There are many ongoing discussions and protocol implementation proposals for the next release of SUSHI for COUNTER Release 4. Many new items will be introduced, both required and optional. Included are new report IDs, identifiers, item data types, categories, and metric types. Because of these changes, the schemas that provide for the delivery of COUNTER reports via SUSHI will need to be updated, and therefore all SUSHI harvesting servers must be updated to be compliant for the next SUSHI release. For SUSHI developers, it is imperative that they stay current regarding the next release of the SUSHI protocol by visiting the Standardized Usage Statistics Harvesting Initiative (SUSHI) website. The SUSHI Developers e-mail list is also an excellent source of information and support.

Along the way, Scholarly iQ plans to continue sharing our knowledge base and observed trending data with the community. We are happy to provide assistance to those wishing to pursue these latest standards. To contact us or to stay current on our recent developments, please visit the Scholarly iQ website or follow us via twitter. We hope that our experiences and observations will be helpful to the community and to anyone who is in the process of or contemplating their own implementations of this pioneering protocol. | IP | doi: 10.3789/isqv23n4.2011.03

GARY VAN OVERBORG <gary.vanoverborg@scholarlyiq.com> is Founder and CEO with Scholarly iQ. **JOHN MILLIGAN** <john.milligan@scholarlyiq.com> is Director of Application Development with Scholarly iQ. **MICHAEL LEE** <michael.lee@scholarlyiq.com> is Lead Data Specialist with Scholarly iQ.

Stay up-to-date on the latest trends by visiting our website or following us on twitter:



www.scholarlyiq.com



twitter.com/#!/scholarlyiq

Euclid Project

docs.google.com/View?docid=d2dhjwd_140d923m7fh&pli=1

Scholarly iQ website

www.scholarlyiq.com

SUSHI E-mail list

To sign up send an e-mail to sushi-developers-subscribe@list.niso.org

SUSHI standard (ANSI/NISO Z39.93)

www.niso.org/standards/z39-93-2007/

SUSHI Report Response Samples

www.niso.org/workrooms/sushi/reports/

SUSHI Schemas and Diagrams

www.niso.org/schemas/sushi

SUSHI Tools & Development Aids

www.niso.org/workrooms/sushi/tools/



RELEVANT
LINKS