

## Case Study



### Record Retention: Web-site Snapshot Spider

Our clients needed to collect and archive snapshots of its public and private websites for audit and compliance purposes. The goal was to create an archived, date-time stamped, fully functional version of their site to answer any questions about content including equity recommendations and market conditions at specific points in time. ebusinessware proposed a solution based around open source tools that the client could launch on an automated basis. The solution had to be easy to maintain and modify.

We believe that similar challenges are faced by a large number of organizations in the financial industry. Following the success of the snapshot-archival solution of public website, ebusinessware also researched, architected and recommended a customized solution based around the same set of open source tools. Based on our implementation experience, we believe that this solution is portable to a variety of web-based, mainframe, client/server-based platforms where electronic records need to be indexed and maintained.

#### Challenges

The client's infrastructure had complexity and breadth. The technical issues that arose during the analysis phase included the following:

- The use of complex JavaScript
- Inclusion of absolute URLs in web links
- Extensive use of frames in the website
- Proprietary naming conventions from an internal client web server
- Archiving pages that require proxy settings along with those that don't

We were faced with the challenge of bringing the solution to a deployable state in a short timeframe. Additionally, the solution had to be

- Simple to use
- Scheduled on automated basis

## Record Retention: Web-site Snapshot Spider

- Easy to maintain - the client did not want high on-going costs
- Quickly deployable

ebusinessware responded to the challenge for rapid time-to-market by forming the team immediately to study appropriate components to the overall solution. To meet the economic constraints, we evaluated an open source product that we customized to meet the specific business requirements. We modified the open source platform to meet the following needs:

- Code modification to handle JavaScript
- Capturing references to absolute URLs and downloading pages separately from the pages referencing them
- Downloading page frames separate from the page itself
- Managing files created in a proprietary format from a third party vendor as separate downloads
- Inventorying sites requiring proxy setting and that downloading them separately with only the requested files
- Our proposed solution also contained the production of large number of metrics data around each site version.

### ***Technical Environment***

We used C / C++ and Microsoft Visual Studio for the development environment.

### **The Team**

- Technical Lead (responsible for interfacing with Business Users and integration testing against business requirements)
- 2 Developers to modify the open source code and to integrate within the client's infrastructure

## Record Retention: Web-site Snapshot Spider

- Subject matter expert and infrastructure advisors.

### The Effort

The Tech. Lead managed all client interaction and development testing, ensuring the project met all of the business requirements. This included weekly status meetings with the business users to review progress and outstanding issues, as well as testing.

- The developers were responsible for the heavy lifting of customizable Open Source Product and unit testing. Separate set of developers provided peer reviews and quality audit check.
- QA setup and testing was performed by a resource from clients.

### The Results

The whole solution was done timely and within the available budget. The entire engagement lasted approximately 4 months from the start of business requirement study to base product selection to development to testing to QA sign-off.

### Conclusion

- The solution required specialized resources across a range of technical and subject matter disciplines
- ebusinessware met the clients needs for low-cost, scalable infrastructure.
- The combination of sophisticated access tools and easily indexed retrieval technology produced a solution that can be tailored to various record acquisitions and retention needs.