



## Web-based Application Development

### Client

The British Library's web curators have a statutory obligation to take regular copies of relevant UK based web sites and archive them for posterity. Due to the absence of any suitable tools to perform this task, The British Library, in conjunction with The National Library of New Zealand commissioned the development of an enterprise class software solution called the Web Curator Tool (WCT), which greatly facilitates a web curator's job. The tool enables the regular scheduling of 'harvests' of nominated sites, the subsequent review and pruning of the harvested content, the submission of approved harvests to the library's archives and the browsing of archived snapshots. With a view to encouraging other national libraries to utilise WCT the source code was placed into the public domain and is maintained on SourceForge.

---

### Assignment Objectives

Because of WCT's position as a core strategic component of the British Library's web archiving infrastructure, the Library engaged PSM's developers to deliver new functionality to further improve the productivity of the web archiving team. The developments were undertaken using the SCRUM agile methodology to provide maximum efficiency in the ongoing development.

The objectives for the assignment were therefore:

- To develop a full understanding of the WCT architecture and functionality
  - To develop a full understanding of the WCT infrastructure requirements
  - To add several new requirements to the solution during a series of 1-2 man month long "sprints" (six sprints have been completed to date)
  - Use the Rally project management tool to manage the iterations and report back to the client
  - To introduced automated daily builds and unit test scripts for regression testing the code
  - To provide ongoing ad-hoc support of the software as it is deployed into production
  - To update the SourceForge project site for WCT to reflect the production of milestone releases
- 

### Key features of the Web Curator Tool

- JSP based Web Client
- Three Java Web applications communicating via Web Services
- Java based web applications hosted by Apache Tomcat, utilising the Spring Framework
- Web crawling functionality is provided by the Heritrix web crawler
- Uses the Hibernate ORM solution to facilitate database agnosticism
- Supported databases are currently PostgreSQL, Oracle and MySQL, but others can be added very easily

In addition to the above functionality, we introduced automated daily builds and unit test scripts to facilitate regression testing.

---

### What we achieved

The enhancements we made to the Web Curator Tool were very well received and our recommended SCRUM based agile development approach was fully adopted by the client. Further development "sprints" have been scheduled for the near future. The same developers have since been commissioned to undertake enhancements to another archiving toolset used by the British Library which is used to preserve electronic copies of scientific journals.