Chapter 4
A Service Component Model and Implementation for Institutional Repositories

Yong Zhang
Tsinghua University, China

Quansong Deng
Tsinghua University, China

Chunxiao Xing
Tsinghua University, China

Yigang Sun
National Library of China, China

Michael Whitney
University of North Carolina Charlotte, USA

ABSTRACT

With the boom of digital resources, there are urgent requirements to set up and manage Institutional Repositories (IRs) for companies and/or organizations. Cloud computing opens a new paradigm to build IRs by providing diverse services. We apply cloud services in the building of IRs and present a new model, which is based on digital object model and service component architecture, and consists of five service components, namely ID, metadata, content, log, and annotation service component. The five components are implemented by five corresponding clouds. These clouds provide two kinds of services: Web service and mashup service. We develop a framework and a code generation tool to generate an IR that can be used to manage the digital resources by invoking the five cloud services. Our approach is applied to the digital library on the history of water conservancy in China of Tsinghua University Library to demonstrate its feasibility.

DOI: 10.4018/978-1-60960-735-7.ch004

1. INTRODUCTION

As institutions and organizations strive to meet the needs of a technologically savvy society, they are confronted with the problem of managing a massive amount of digital resources. If they are to meet societal needs, they must incorporate a digital resource management system that is capable of affording individuals an opportunity to access their digital resources in a readily available and useful manner. In response, many have started to use an Institutional Repository (IR) which is an online system for digitally collecting, preserving, and disseminating intellectual output (Smith 2002) (Institutional Repository 2010). Although IR originates from academic fields, more and more organizations are setting up their own IRs that extend the original academic elements such as institutionally defined, scholarly content, cumulative and perpetual, interoperability and open access (Johnson 2002).

Meanwhile, Cloud Computing has been booming for several years. It is an extension of Grid Computing which emphasizes the on-demand usage of computing powers. In comparison, Cloud Computing focuses on higher-level services or well-designed services. More so, Clouds are defined as a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services) (Vaquero et al. 2009). These resources can be dynamically configured to adapt to variable load (scale), leading to optimum resource utilization. This pool of resources is typically utilized on a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs.

Overall, Cloud Computing has three advantages that can help us manage digital resources. First of all, it is a pay-per-use utility model that decreases the cost of setting up, operating and upgrading an Institutional Repository. Secondly, it is scalable to deal with the massive digital resources that are continuously growing. Thirdly, it provides flexible services to be integrated with existing applications.

In addition, there exist many legacy digital resource management systems that require integration with the emerging systems. The current software usually provides Web Services to be invoked from the old systems. However, to integrate these Web Services, it is necessary to modify the codes and re-compile which can be very cumbersome. So as to improve upon this practice, we offer Mashup services as a means to support this integration. This is because Mashup Service can be integrated at a higher level, which reduces the workload to some extent. It serves as a straightforward way to create new Web applications by combining existing Web resources utilizing data and Web APIs (Benslimane et al. 2008).

The software that we have developed is the Digital Resource Management System Version 2.0 (DRMS2.0), which is an upgrade of the DRMS1.0 which was a stand-alone system that provided the management of digital resources from ingestion, storage, index to services. For DRMS2.0, we design it as a middleware between application layer and network layer, which includes five clouds, a digital resource management container and an assistant tool. The five clouds provide three kinds of services: SaaS, Mashup Services and Web services. The SaaS services include management of digital object identifiers and digital content. Mashup service consists of Web APIs to be integrated in the user interface of a Web application. Web services provide the programmable interface to operate the digital resources. The digital resource management container called MenuFrame provides the Web interfaces, role-user management, privilege management, and supports central authentication. The assistant tool helps the user create an Institutional Repository for the customized digital resources, or just an empty framework for common use. The five clouds themselves are also created by this tool.

In traditional digital resource management systems, the content of the digital resource is normally static after it is ingested. But from our