Moving Digital Library Service Systems to the Grid

Leonardo Candela¹,², Donatella Castelli¹, Pasquale Pagano¹,², and Manuele Simi¹

¹ Istituto di Scienza e Tecnologie dell’Informazione “Alessandro Faedo”,
Consiglio Nazionale delle Ricerche,
Via G. Moruzzi, 1 - 56124 PISA - Italy
{candela, castelli, pagano, simi}@isti.cnr.it
² Università degli Studi di Pisa, Dipartimento di Ingegneria dell’Informazione,
Via G. Caruso, 1 - 56122 PISA - Italy
{candela, pagano}@iet.unipi.it

Abstract. The architecture of a digital library service system strongly influences its capabilities. In this paper we report on our experience with a distributed digital library system, OpenDLib, and we describe why, in the attempt to better satisfy the user requirements, we decided to develop DILIGENT, a service-oriented digital library infrastructure on the Grid. The paper describes and compare the two systems by focusing, in particular, on that part of the architecture that controls and supplies the necessary features for creating and managing digital libraries.

1 Introduction

Four years ago, the DLib group at ISTI-CNR began to develop a Digital Library Service System (DLSS), i.e. a system for creating and managing digital libraries (DLs). When designing this system, named OpenDLib [8], our goal was to create a customizable system that, if appropriately configured, could satisfy the needs of different application frameworks.

Our first step was to clarify what we meant for DLs and to identify the features that a system able to implement DLs should provide. This analysis brought us to understand that a DLSS is a complex system that must not only offer powerful user functionalities (e.g. search, browse, annotation) but also implement basic functions for supporting the fruition of the user functionality and for guaranteeing the quality of the overall DL service, e.g. its availability, scalability, performance. Moreover, it must satisfy a number of other desiderata, like being extensible, easy to install and to maintain.

In order to create the conditions for achieving the required level of quality we analyzed a range of possible system architectures and, finally, we decided to adopt a distributed, dynamically configurable, service-oriented architecture (SOA). The implementation effort spent in developing a DL system based on this architecture turned out to be much greater than that required for implementing a centralized one since a number of services dedicated to the co-ordination, management and optimal allocation of the different service instances had also to be provided. OpenDLib is now an operational system that has been used for building a number of DLs [21–24]. Each of these DLs has its own specific distributed architectural configuration that reflects the needs of the application scenario where it operates. In all these different experiences the distributed
service architecture has proved to be a valid instrument to satisfy a number of requirements that could not have been met otherwise. The greater development effort has thus been highly compensated by the better quality and organization of the DL functionality exposed to the users.

New architectural approaches have emerged, or have been consolidated since we designed the OpenDLib system, e.g. Web services [5, 10, 20], P2P [19], Grids [15, 16]. All these approaches provide features that simplify the realization of a distributed DL architecture by offering a standardized means of building software services that can be accessed, shared, and reused across a network.

The DLib group at ISTI-CNR, with a number of other European research organizations and software companies, has recently initiated a new project, A test-bed Digital Library Infrastructure on Grid ENabled Technology (DILIGENT), which, by exploiting these new approaches, will develop an infrastructure for supporting a new method for the creation and operation of DLs. By using this infrastructure user communities will be allowed to create multiple on-demand transient DLs active on the same set of shared resources. This infrastructure will have a new architecture which integrates the service-oriented approach with the middleware provided by EGEE [12], the project that will deliver the largest European Grid Infrastructure.

In this paper we introduce the OpenDLib architecture by focusing on the aspects related to the services management and we report on our experience in operating this system. Then, we describe why we decided to move towards the new DILIGENT infrastructure, and present its architecture and the services that are responsible of the dynamic creation and optimal handling of the DLs.

The rest of this paper is structured as follows: Section 2 introduces the main requirements that motivated our choice of a distributed service architecture; Section 3 illustrates the main functional areas of an architecture for DLSSs; Section 4 describes OpenDLib and, in particular, its architecture in terms of the identified areas; Section 5 discusses the motivation that brought to the introduction of the new DILIGENT infrastructure; Section 6 describes the architecture of this infrastructure and, finally, Section 7 concludes.

2 Requirements for a Digital Library System

We have been working on DL applications since 1995. During this period we have met several potential user communities and have discussed their requirements with them. Some of these requirements have strongly influenced the architectural design of both OpenDLib and DILIGENT. In particular:

1. There is a set of core DL functionalities, such as search, retrieval, access to information objects, that any DL should provide. The format in which each of these functionalities is presented to the user is usually different since it complies with the application specific vocabularies and rules. In addition to the core functionalities, each DL, usually, must provide other specific functionalities for serving application-specific requirements.