

# Abstract

There is an increasing interest on digital archives for the preservation of digital documents. The main software infrastructure of a digital archive is its repository, which stores the digital objects. Repositories can be created by deploying repository systems.

This project was run in the context of Gisa, a software application developed by ParadigmaXis for the management and description of the content of archives. Since in digital archives the content is digital, it is possible to couple it more closely with software applications. The objective of this project was to perform a comparative study on repository systems, and to develop an extensible module for the integration of Gisa (and other applications) with repositories systems.

In the first phase, the comparative study was performed on three open source repository systems: Fedora; DSpace; and EPrints, over the aspects: interoperability; distribution; scalability; design; security; and UI configuration. Additionally, the commercial product DigiTool was included in the interoperability section of the study, which listed the exposed APIs and standard protocols implemented per functionality. This study set the stage for the development of the interoperability framework. It is based on a hierarchy of adapters, with a common generic interface to the contents of repositories, from which less abstract adapters are derived to implement protocols to access these contents in specific repositories. The framework was implemented as a C# library to be used by applications, and includes a univocal and homogenous identifier scheme and search functionalities that allow a query to be run on multiple repositories simultaneously. Adapters were implemented for repositories exposing OAI-PMH; for repositories based on Fedora; and for simple files viewed as a repository. Applications were developed using the framework and include an OAI-PMH harvester and a repository browser, which uses the generic interface of the framework and can therefore be used to browse any repository with the correspondent adapter.

# Resumo

Existe um aumento no interesse em arquivos digitais na preservação de documentos digitais. A principal infraestrutura de software de um arquivo digital é o repositório, que armazena objectos digitais. Repositórios podem ser criados com base em sistemas para repositórios.

Este projecto decorreu no âmbito do Gisa, uma aplicação de software desenvolvida pela ParadigmaXis para a gestão e descrição dos conteúdos de arquivos. Como em arquivos digitais o conteúdo é por definição digital, é possível ligar este mais estreitamente a outras aplicações de software. O objectivo deste projecto foi efectuar um estudo comparativo de sistemas para repositórios, e desenvolver um módulo para a integração de repositórios com o Gisa e outras aplicações.

Na primeira fase foi feito o estudo comparativo em três sistemas *open source* para repositórios: Fedora; DSpace; e EPrints, sobre os aspectos de interoperabilidade; de distribuição; escalabilidade; *design*; segurança; e configuração do UI. Um produto comercial, DigiTool, foi também incluído no estudo, na secção de interoperabilidade, que lista por funcionalidades as APIs expostas e protocolos normalizados implementados. Este estudo criou as condições para o desenvolvimento de uma *framework* de interoperabilidade. Esta é baseada numa hierarquia de adaptadores, com uma interface genérica comum, da qual são derivados adaptadores menos abstractos que implementam protocolos para o acesso aos conteúdos de um repositório específico. A *framework* foi implementada como uma biblioteca de C# para ser usada por aplicações, e inclui um esquema de identificação unívoca de objectos digitais e integra funcionalidades de pesquisa que permitem interrograr vários repositórios simultaneamente. Foram implementados adaptadores para repositórios com OAI-PMH; para repositórios baseados em Fedora; e para ficheiros simples vistos como sendo um repositório. Foram desenvolvidas aplicações usando a *framework*, incluindo um *harvester* para OAI-PMH e um *browser* de repositórios que utiliza a interface genérica e portanto pode ser usada com qualquer repositório com um adaptador.