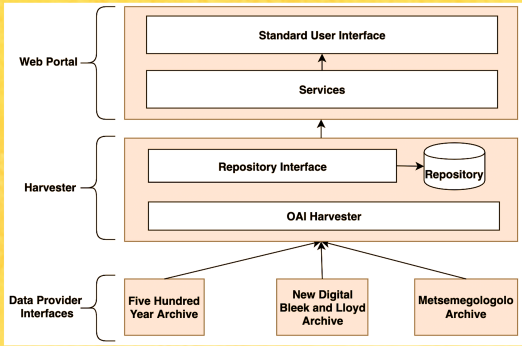
HERIPORT

A Low-Cost South African National Heritage Web Portal built with Metadata Aggregation

Project Objectives

This project aimed to create the first national low-cost heritage portal that provides access to multiple local archives, using Metadata Harvesting via the Open Archive Interface Protocol for Metadata Harvesting (OAI-PMH). The system was divided into three components: The Data Provider Interfaces, The Harvester and the Web Portal with search and browse services.



Data Provider Interfaces

The Data Provider Interfaces map the original metadata from the Data Providers to (unqualified) Dublin Core (DC) and expose the metadata using the OAI-PMH. Three DC Converter applications were implemented to map the original metadata to DC and three OAI-PMH Data Provider Interfaces were developed to expose the DC metadata to be aggregated by the Harvester. Each interface provides OAI-PMH request and response services to view the metadata of a repository.

Harvester

The Harvester aggregates data centrally from multiple resources, conforming to the OAI-PMH. The metadata is retrieved from the Data Provider Interfaces and is subsequently converted and indexed into Apache Solr, which provides an interface for the Web Portal. The component can easily be scheduled to run daily or weekly using scheduling packages such as Cron in a Linux environment.

Web Portal

The Web Portal provides a standard user interface for end-users to access the aggregated resources. The harvested data is retrieved from the Apache Solr Server and presented to the end-user, providing the services: search, browse and a carousel feature. Users are able to browse the resources by title, archive and date, and can search using keywords from all fields. They are presented with DC metadata for resources and are able to be redirected to the original resource via an identifier.

Project Results and Conclusions

Following respective testing:

- The Data Provider Interfaces could convert the metadata to DC and expose the metadata using the OAI-PMH as an integrable and reusable component.
- The Harvester retrieved, converted and indexed metadata records in an efficient and interoperable manner that provided for performance and scalability.
- The Web Portal was able to provide cross-archive discovery services, maintaining usability and efficiency.

Overall, the HERIPORT system successfully implemented three separable and reusable components.

University of Cape Town
Department of Cape Town
Email: dept@cs.uct.ac.za
Tel: 021 650 2663

Project Team
Alex Priscu (Data Provider Interfaces)
Ashil Ramjee (Harvester)
Toshka Coleman (Web Portal)
Supervised by Dr Suleman