



# Project Gandhara

**Shedding Light on the Hidden Web**

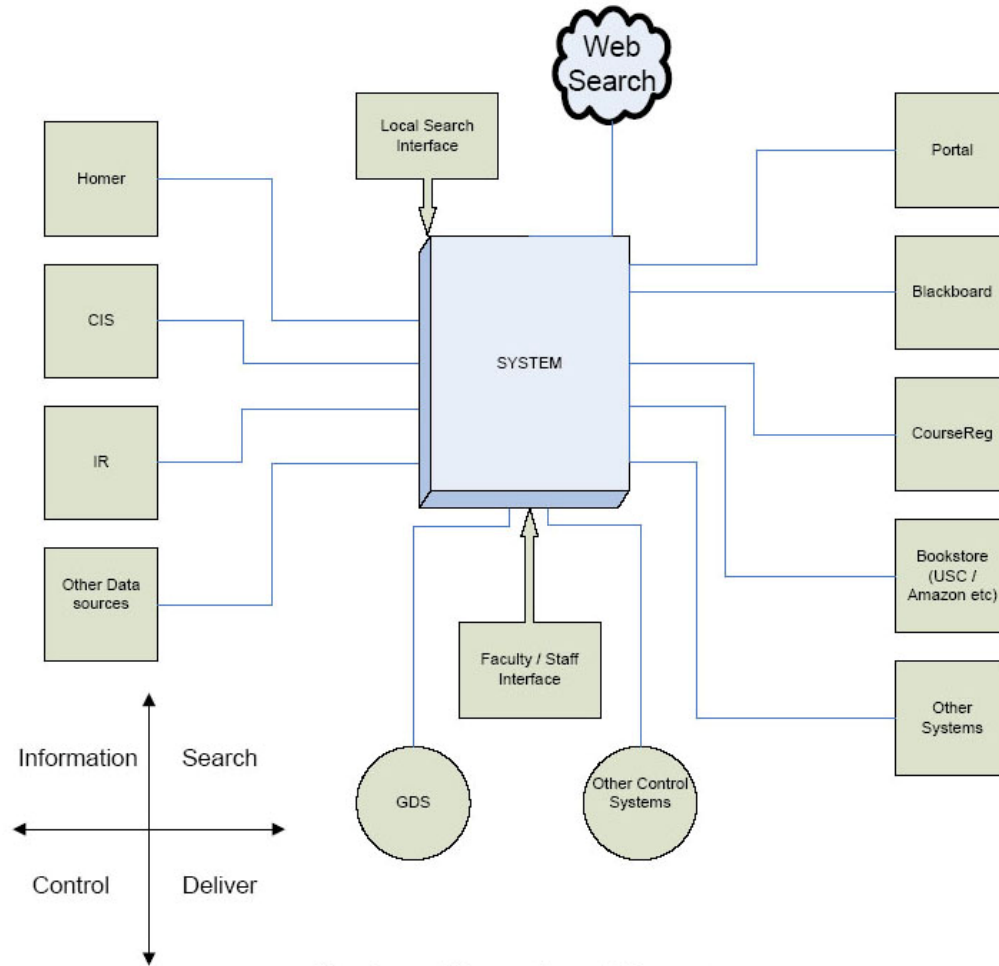
**PRDLA/PNC Annual Meeting 2005**

**Honolulu Hawai'i**

## *Statement of Purpose*

- Multiple Data Sources and Formats
  - Library collections
  - USC research data
- Multi-format Data Ingest
  - Descriptive Metadata Elements
- Indexable Output
  - XML or XHTML
  - Both

### System Architecture



System Overview Diagram

## *MARC/OAI Ingest and Transformation*

- Catalog Data
  - Standard library catalog: SIRSI
  - 2 Stage Process
    - MARC – MARCXML
    - MARCXML – XHTML
- OAI Data Ingest
  - Descriptive metadata elements

## *Indexing and Extraction*

- Local Indexing
  - Lucene
    - Configure display
  - Verity
- Remote Indexing
  - Google et al.
- Content Delivery
  - Custom web page or news feeds

## PROJECT Gandhara

Results: 1 - 10 of 45

**An account of the present state of the Hebrides a ... [Homer Record]**

An account of the present state of the Hebrides and western coasts of Scotland in which an attempt is made to explain the circumstances that have hitherto repressed the industry of the natives; ... being the substance of a report to the Lords of ...

Author: **Anderson, James,**

**Portrait of the office staff of the Western Union ... [Digital Archive Record]**

Portrait of the office staff of the Western Union Telegraph Company, 1891. Photographic portrait of the office staff of the Western Union Telegraph Company, 1891. There are eighteen people in the picture, seven women and eleven men, and they are ...

Subject: **Telegraph Employees Western Union Telegraph Company Portraits -- Group -- Occupations -- General**

## *Next Steps*

- Data Mining
  - Disparate data sources can be amalgamated, indexed and searched
- Semantics
  - Taxonomies, terminologies, and controlled vocabularies (LCSH, MESH)



# Demo

**Todd Grappone**

Associate Executive Director, Information  
Development and Management

Information Services Division, Resources and  
Services

University of Southern California